UNIVERSITY OF LIVERPOOL

Exploring the role of phonological and motor learning in word and phrase frequency effects in speech production: The effect of word frequency, phrase frequency and phonological complexity on speech production in children and adults

By Rachel Quinn

Thesis submitted in accordance with the requirements of the University of Liverpool for the degree of Doctor in Philosophy

Department of Psychology

University of Liverpool

January 2021

UNIVERSITY OF LIVERPOOL

INSTITUTE OF LIFE AND HUMAN SCIENCES

PSYCHOLOGY (SCIENCE)

DECLARATION IN HIGHER DEGREE THESES

DECLARATION

This thesis is the result of my own work. The material contained in the thesis has not been presented, nor is currently being presented, either wholly or in part for any other degree or qualification.

Signed .... (Candidate)

Signed 14/01/2021 .... (Candidate)

Acknowledgements

Firstly, I would like to thank Dr Colin Bannard for the invaluable advice and continuous support during my PhD. I really appreciate all the time you have given me. I would also like to thank Professor Julian Pine for his insight and comments. Without their expertise and input, none of the work in this thesis would have been possible.

I am extremely grateful to the schools, the children who took part in my research and the teachers for their time. I would also like to express gratitude to the Department of Psychology for employing me and funding my PhD.

To my friends and family, thank you for being so fun. Davide, thank you for being so funny. You are all a constant source of support and I am very lucky to have you in my life.

Abstract

The objective of this thesis is to explore the effect of phrase frequency, word frequency and phonetic complexity (and their interactions) on production fluency in both children and adults. We explore firstly whether the previous findings for independent effects of phrase frequency, word frequency and word complexity on phrase production hold when using phrases that are simultaneously manipulated for all three factors, and secondly how the effects of these properties interact. In particular we aim to explore whether children's phonological learning, like that of adults, involves phonological/motor chunking. This refers to the idea that frequent sequences are represented in phonological working memory as a single unit which can be accessed directly, rather than having to be formed on the fly during production. Therefore, words that are difficult to say when first encountered become easier to articulate due to practice. An interaction between phrase/word frequency and complexity, where complex items are produced more fluently when they are also more frequent, would indicate that phonological/motor chunking is taking place. We also test the secondary prediction that phrase and word frequency will affect (promote or suppress) each other’s effect on accuracy.

In chapter 3 and 4 we asked children to repeat phrases manipulated for phrase frequency, word frequency and complexity, while chapter 5 uses the same stimuli in adults. We found that, overall, children do use phonological/motor chunking, but only for phrases where they have robust conceptual and phonological knowledge. Furthermore, we found that this effect was driven solely by word frequency in our stimuli - there was no effect of phrase frequency on production in children. Using the same stimuli, we found that adults also showed evidence of phonological/motor chunking for words which were high in frequency, although we found evidence of competition between word frequency and phrase frequency effects that was not seen in the children. Finally, Chapter 6 reports on 2 experiments which use a training session to manipulate non-word sequences for phrase frequency and word frequency. Again, we found that word frequency drives phonological/motor chunking during production.

The results from these experiments suggest that young children and adults do undergo phonological/motor chunking. However, importantly, this effect is driven by word frequency - the phrase frequency effects seen elsewhere in the literature must be attributed to other mechanisms (e.g. long term memory associations between the constituent lemmas contained in high frequency phrases).

List of Figures

List of Tables

1 Introduction to the Thesis and Background Literature

1.1 Statement of the problem

The starting point for this thesis is the known effect on production fluency for multiword speech (particularly in young children) of three linguistic properties – the phonological complexity of the words they are producing, the frequency in the language of the words that they are producing, and the frequency in the language of the word sequence they are producing. It has regularly been shown that speakers take less time and make fewer errors when producing words that are less phonologically complex and words that are more frequent in language. In recent years, a similar advantage for phrases that are more frequent in language has been found. However, the existing literature only considers the effects of these factors in isolation. In real language use they are intertwined and are often confounded (less frequent words and phrases tend to be less phonologically complex, frequent phrases tend to be composed of frequent phrases, etc.). The goal of this thesis is first to explore whether the previous findings for independent effects of phrase frequency, word frequency and word complexity on phrase production hold when using phrases that are simultaneously manipulated for all three factors, and second to understand how the effects of these properties interact. As will be explained, the question of whether and how these factors interact has implications for both how we should interpret previous findings, as well as for models of speech learning and development more generally.

1.2 Background literature

We are interested in why speakers (and particularly children) find it easier to produce some words/phrases compared to others. This chapter reviews the existing literatures relevant to three predictors of fluency - word frequency, phrase frequency and phonological complexity. Although it is generally accepted that both children and adults are better at producing words/phrases which are frequent in language compared to words/phrases which are infrequent, the processes underlying these effects are less clear. We are also interested in how these frequency effects might interact with each other, and with complexity. Theoretically, less complex words should be easier for speakers to produce. However, we must learn to produce language in a way which is most useful for us to communicate. Therefore, speakers learn to produce the words they will use most often, regardless of complexity. In the following chapter, we aim to explore the underlying theoretical frameworks which consider these frequency effects. We will consider different models of production in terms of both frequency and complexity.

1.3 Word frequency

There is a great deal of published evidence from both comprehension and production tasks which show that speakers are sensitive to the frequency in the language of the forms they are producing. Many comprehension studies have found a processing advantage for high frequency words. For example, speakers will use high frequency words more quickly, and with fewer errors, than low frequency words in reading and auditory word recognition tasks, while high frequency words are judged to be real words more quickly during lexical decision tasks (see Ellis, 2002 for a review). Similarly, production tasks demonstrate that speakers are more efficient in producing words that are more frequent compared to words that are less frequent. Consistent with the focus of this thesis, the following will discuss the reported effects of word frequency on different production tasks for both children and adults. Given the scale of the word comprehension literature, and its marginal relevance to our questions, this literature will not be considered further in this chapter.

1.3.1 Word frequency effects in adults

Word frequency has been shown to predict fluency in adults in a range of experimental paradigms. One method used to explore the effect of different linguistic factors on production is picture naming. This involves a speaker being presented with a picture which depicts a target word or sequence. The speaker is asked to name what they see, and the subsequent production is analysed - usually for the duration of production or error rate. An early picture naming task conducted by Oldfield and Wingfield (1965) was one of the first to show that adult speakers are faster at producing frequent names (e.g., boat) in comparison to infrequent names (e.g., broom).

Many studies have indirectly provided evidence for word frequency effects using picture naming paradigms, while focused primarily on other manipulations (e.g., exploring the level at which frequency effects occur during lexical access – see section 1.6.2). Strijkers, Costa and Thierry (2010) aimed to explore the electrophysiological signature of lexical access in speech production, where both the naming latencies and latencies of event-related potential experiments were measured during a picture naming task. In terms of naming latency, it was found that Spanish-Catalan bilingual speakers are faster to name pictures with high-frequency names compared to low frequency names.

To test the predictions of different models of production, Janssen, Bi and Caramazza (2008) looked at whether the production of compounds (a single word made up of 2 or more words e.g., windmill) is affected by morphemes or the compound's whole-word frequency. The authors used a picture naming task to measure speed of production for compound words and found that high whole-word frequency of compounds elicited faster naming times in both Mandarin Chinese and English speakers.

Italian speakers produce high frequency gender marked verbs more quickly than their low frequency equivalents in a picture naming task where the speaker was required to produce gender-marked verbs plus their pronominal clitic utterances (Finocchiaro & Caramazza, 2006), while Spanish speakers who performed a gender decision task during picture naming were faster when naming sequences containing high frequency nouns compared to sequences containing low frequency nouns (Navarrete, Basagni, Alario & Costa, 2006). Dutch speakers are faster at producing names for high frequency pictures compared to low frequency pictures even when the target words are controlled for word length and morphological complexity (Jescheniak & Levelt, 1994) and high frequency words in both semantic homogeneous and semantic heterogeneous contexts are named more quickly than their low frequency comparison (Santesteban, Costa, Pontin & Navarrete, 2006).

Another experimental paradigm used to explore the effect of word frequency on production is a word naming task. Here, the speaker is presented with a target word, normally in written form, and is asked to produce what they see. Dell (1990) used an error inducing paradigm where adults were visually presented with phrases containing homophones manipulated for word frequency (e.g., 'by the pin' vs 'buy the pin'). The target phrases were selected so that they were likely to generate errors (e.g., exchange error 'pie the bin') and error rates for the productions were measured. It was found that less phonological errors occur in more frequent words for both function and content items.

Word frequency has also been shown to predict fluency in studies focusing on naturalistic speech. Speakers will hesitate more often for content words which are low in word frequency or low in contextual probability (Beattie and Butterworth, 1979). Low-frequency inflected forms have a higher rate of no-marking errors (e.g., producing 'walk' instead of 'walked') compared to high-frequency forms (Stemberger & MacWhinney, 1986). Substitution errors occur more often in low frequency words compared to high frequency words (Stemberger, 1984) and more phonological errors occur in naturalistic speech for less frequent words (Harley & MacAndrew, 2001), though more complex (longer) words were also found to elicit more errors.

The effect of frequency on production fluency has also been reported for various clinical populations using these paradigms. Aphasic patients show an advantage for producing high frequency words. In a series of picture naming tasks, Nickels (1995) found that patients made more phonological, but not semantic, errors for words which were low in frequency compared to high in frequency, while Kittredge, Dell, Verkuilen and Schwartz (2008) found that patients with aphasia made fewer phonological and semantic errors in more frequent words than in less frequent words. Knobel, Finkbeiner and Caramazza (2008) found a patient with aphasia made fewer errors for non-words which were manipulated to be more frequent in comparison to non-word which were manipulated to be less frequent. Although the error types varied as an effect of lexical frequency, the overall frequency effect held when effects of imageability, length, and number of phonological neighbours were controlled for.

Finally, elicited imitation tasks have been used to explore the effect of word frequency on fluency in adults. For this task, participants listen to and are then asked to repeat specific target stimuli (normally a word or sequence of words), where the subsequent production is analysed. Vitevich and Sommers (2003) found that when adult speakers listened to definitions of words which varied in frequency, more tip of the tongue states were elicited for target words with low frequency, compared to target words with high frequency.

1.3.2 Word frequency effects in children

Word frequency effects are also well-substantiated in studies which focus on production in children for many of the above reported methods. For instance, there is a clear effect of word frequency during naturalistic speech. Ota and Green (2013) analysed the production of word-initial consonant clusters in the spontaneous speech of 3 children aged 0;11–4;0 and found that the maternal input frequency of a given word predicted the age at which cluster production accuracy first reached 80% for this word. Furthermore, this effect held when controlling for production frequency. However, this effect was found to vary by type of cluster, where most cluster types acquired earlier in production were found to be less affected by input frequency. Using the same corpus, Jones (2020) explored the accuracy and variability of spontaneous word productions of 5 typically developing American English-speaking children aged 11 months to 4 years old, and found that high frequency words (in child directed speech) are produced more accurately than low frequency words.

As for adults, effects of frequency on production have also been reported in clinical child populations. Studies which focus on predictors of stuttering also provide evidence for a word frequency effect, where children are more likely to stutter for words which are less frequent. Anderson (2007) found that children were more likely to experience stuttering-like disfluency for words that were lower in frequency in comparison to a fluently produced control word. Similarly, both typically developing children and children who experience word-finding difficulties (aged 7 to 12 years old) were shown to be better at producing high frequency words compared to low frequency words in an oral naming task (Newman & German, 2002). This effect held over all age groups, though Newman and German (2005) found a larger effect of word frequency in typically developing adolescents than typically developing adults when using a similar oral naming task on speakers aged 12 to 83 years old.

Elicited imitation paradigms have also been used to explore the effect of frequency on fluency in children. Sosa and Stoel-Gammon (2012) used an elicited imitation paradigm to explore the effect of word frequency on production in children aged 2;0-2;5. In this case, the researcher elicited productions of the target words during play, naming and reading. The authors found that the children's productions of high frequency words were less variable than those of low frequency words, though more complex words (containing later developing consonants and syllable structures) were also produced more variably.

Elicited imitation paradigms have also been implemented for artificial stimuli. Rather than studying the production of real words of differing frequency, other researchers have sought to achieve greater control over their stimuli by conducting training studies with artificially created word-like sequences. One possibility is that the word frequency effect is a result of motor learning, where words that are difficult to say when first encountered become easier to articulate as a result of practice of the motor routine (more will be said about this when we come to consider models of production later in this chapter – see section 1.6). However, it is also possible that this frequency effect is driven by the frequency-related differences in the meaning of words. By creating artificial stimuli researchers have sought to separate motor from semantic factors.

Sasisekaran, Smith, Sadagopan and Weber-Fox (2010) asked children aged 9–10 years and adults to produce non-words over 2 sessions in an elicited imitation experiment. Both groups were faster and made fewer errors when producing non-words after practice. In this case, adults produced the most complex non-words more accurately and more quickly on the second day compared to the first day, while the children's production of the least-complex non-words improved on the second day compared to the first day.

A similar study looking at the trade-off between phonological complexity and frequency using an elicited imitation task was reported by Segawa, Tourville, Beal and Guenther (2015). Mono-syllabic non-words containing phonotactically legal (*e.g., BLERK, THRIMF, TRALP*) and illegal (*e.g., FPESHCH, GVAZF, TPIFF*) bi- or triconsonantal initial and final consonant clusters showed that speakers had a significantly lower error rate after practice for the illegal sequences but did not differ in their repetition of legal syllables. This finding, where speakers only improved in their repetition of illegal syllables, suggests that speakers have undergone motor learning. Speakers, it is proposed, improved in their repetition of illegal consonants but did not improve in their productions of legal consonants because they already had learned motor routines for these legal consonant clusters. Furthermore, after introducing novel illegal sequences, speakers made more errors during novel illegal syllables compared to learned illegal syllables, which provides further evidence for motor learning. This latter finding shows that the more efficient repetition of illegal sequences after practice was not merely a result of learned phonotactic rules because performance improvements were specific to the stimuli learned. If improvements were a result of learning of rules, then the novel illegal syllables that followed the same phonotactic rules would have also improved. Instead, this improvement is specifically due to motor learning.

Therefore, it is plausible that motor learning, arising from practice, is a component that contributes to frequency effects. This will be considered in more detail in section 1.6, where word frequency will be considered in terms of models of production.

1.4 Phrase frequency

In addition to the word frequency effects discussed above, there is also evidence to suggest that speakers are sensitive to the frequency of multi-word sequences in language. An early study conducted by Bybee and Scheibman (1999) found that speakers are more likely to reduce their production of ‘don’t’ during naturally occurring speech when ‘don’t’ is contained in a highly frequent sequence, such as ‘I don’t know’ or ‘I don’t think’. Similarly, Bell et al (2003) found that the production of function words (e.g., ‘you’) contained in frequent collocations (words that often co-occur e.g., ‘you know’) are shorter in duration and have less-full forms (e.g., reduced vowel) compared to less frequent function words in naturalistic speech. These early findings led to speculation that there are lower processing demands for more frequent sequences. Therefore, the following will discuss the reported effects of phrase frequency in both comprehension and production tasks, while also considering how multi-word phrases are processed and represented in the lexicon.

1.4.1 Comprehension tasks

We discuss the effect of phrase frequency in the comprehension literature to show the scope with which learning is driven by experience. However, the underlying mechanisms which are involved during comprehension tasks differ to that of production tasks (in this thesis we use production tasks to explore our research questions). The comprehension literature generally reports on experiments which use input frequency – this is the frequency with which words/phrases are heard/understood by the speaker. While the production literature generally reports on experiments which use production frequency – this is the frequency with which words/phrases are retrieved/produced by the speaker. This is relevant because the underlying processes which drive input frequency effects likely differ to that of production frequency effects.

Models of production generally accept that there is a semantically/syntactically specified representation (sometimes called conceptualisation) followed by a phonologically specified representation (sometimes called formulation) and this system is said to be organised by frequency, with more frequent words/sounds being more easily accessed (see section 1.6). If we accept this model, for production tasks, words/phrases which are retrieved more often (words/phrase with high production frequency) will show an advantage during speech. While for comprehension tasks, the observed frequency effects are driven by input frequency during the conceptualisation level (where speakers have more robust conceptual knowledge of words that they have heard more often). Therefore, the following explores the effect of phrase frequency on fluency using comprehension tasks, with consideration to the underlying process of phrase frequency for comprehension.

Evidence for reduced processing of phrases which are high in frequency have been well established in comprehension tasks. One method which has been used to measure this is self-paced reading tasks, where the speed with which speakers read target stimuli is measured. During self-paced reading, processing speeds have been shown to reflect the frequency of phrases, with more frequent phrases being read more rapidly than infrequent phrases.

Bannard (2006) used a self-paced reading task to measure the processing speeds for different phrases manipulated for frequency. Adult speakers were visually presented with frequent phrases (e.g., *a state of emergency*) matched to infrequent phrases (e.g., *a state of pregnancy*), which were identical except for the final word - which had a similar frequency to its matched phrase. Here, subjects were found to be faster when reading frequent phrases compared to infrequent phrases. Furthermore, this effect held in a second experiment where speakers were presented with frequent and infrequent phrases identical in terms of length and syntactic form (e.g., frequent sequence ‘on the other side of’ was matched to infrequent sequence ‘on the great slab of’).

Similarly, Shantz (2016) used a self-paced reading task for matched sets of high and low frequency 4-word sequences manipulated for grammaticality and found that both native and second language speakers took less time to read high frequency phrases in grammatical and ungrammatical conditions. Furthermore, Reali and Christiansen (2007) show pronominal object relative clauses are read more quickly when they begin with a high frequency pronoun–verb compared to a low frequency pronoun–verb sequence and Tremblay, Derwing, Libben and Westbury (2011) found that sentences containing lexical bundles (e.g., *in the middle of the…*) were read more quickly than their control sentence pair (e.g., *in the front of the…*).

Eye-tracking has also been used to measure self-paced reading. Siyanova-Chanturia, Conklin and van Heuven (2011) used binomials (phrases containing 2 content words linked by a conjunction, where a specific word order is more frequent than the other e.g., ‘bride and groom’ compared to ‘*groom and bride*’) to explore the effect of phrase frequency on reading times. Both native and non-native speakers processed frequent binomials more quickly compared to less frequent binominals. Furthermore, when both groups of speakers were shown binomials in their frequent and reversed forms, it was found that both native and advanced non-native speakers read the binomial phrases more quickly than their reversed less frequent equivalent.

Underwood, Schmitt & Galpin (2004) used idioms in an eye-tracking self-paced reading task for native and non-native speakers. Idioms refer to frequent multi-word sequence ‘sayings’ which often contain a figurative meaning (e.g., ‘*at the end of the day*’). Here speakers were presented with an idiom (e.g., ‘*honesty is the best policy…*’) and novel sentence containing same lexical item (e.g., *'it seems that his policy of…*'). Native speakers had fewer and shorter fixations for idioms compared to their novel phrase equivalent, whereas there was no effect in the non-native group. This suggests that this frequency effect might emerge later in development.

Similarly, native and non-native speakers who read a series of stories containing an idiom used figurately ('*at the end of the day'*), literally ('*at the end of the day*' – in the evening), and in a novel sentence ('*at the end of the war*') showed native speakers were faster to process figurative and literal meaning than novel phrases. However, non-natives read idioms and compositional phrases at the same speed and processed figurative uses of idioms more slowly than literal ones. The authors argue that this latter result is due to context instead of frequency (Siyanova-Chanturia, Conklin & Schmitt, 2011). However, it is also possible that these findings are driven in part by semantic compositionality.

These reported phrase frequency effects raise a problem for the words and rules approach to language (e.g., Pinker and Ullman, 2002). This theory posits that there is a clear distinction between the lexicon, an inventory of memorised word forms, and grammar. Sentences are constructed by combining memorised word forms taken from the lexicon via knowledge of innate rules (e.g., the memorised word form '*walk*' might be combined with the regular inflection rule, where 'ed' is added to create the past tense verb '*walked*'). This model predicts that frequency will affect the processing of word forms in the lexicon but will not affect the processing of computed multi-word sequences. However, given that phrase frequency effects mirror word frequency effects, it is possible that both words and phrases are represented and processed in the same way. One explanation as to why speakers have a reduced processing speed for high frequency phrases is that frequent sequences of words are represented together as a single unit in the lexicon. In this case, frequent phrases are stored in long term memory to compensate for the limited resource of working memory. A reduced working memory load is then required to access the phrase. However, it is less clear whether these units are holistic (which means that phrasal representations are unrelated to any representations for the component words) or whether the representations for the component words are also accessed during production.

Reaction times on decision and discrimination tasks have also been used to measure frequency effects in comprehension. Arnon and Snider (2010) used a phrasal-decision task to measure processing latencies in speakers who were presented with 4-word expressions manipulated for frequency (e.g., *don't have to worry* vs *don't have to wait*), where the final word, the bigram and the trigram were controlled for. It was found that speakers were faster to respond to high frequency phrases compared to low frequency phrases across the frequency range.

Sosa and MacFarlane (2002) used a word monitoring task to measure adult reaction times to the word 'of' for collocations (e.g., '*part of*') which varied in frequency. It was found that response latencies increased when the collocation was higher in frequency. The authors took this result as evidence of competition between the representation of the word and the phrase. Reaction times were faster for phrases which were more predictable but were slower when there was competition between the particle and 'chunked' collocation. This is taken as evidence of holistic processing. Since reaction times were longer for high frequency collocations, it is possible that access to 'of' was hindered given that the phrase is stored as a holistic sequence. Because the phrase is autonomous from the individual subparts of the phrase, 'of' on its own was difficult for speakers to access.

In recognition memory, low frequency words are remembered more easily than high frequency words. This is known as the mirror effect, where more unusual words stand out to the speaker. Jacobs, Dell, Benjamin and Bannard (2016) aimed to explore whether the mirror effect would stand for phrases. However, it was found that previously seen high frequency phrases were discriminated from previously unseen high frequency phrases as accurately as previously seen low frequency phrases were discriminated from previously unseen low frequency phrases. This suggests that phrases are not stored holistically. Furthermore, phrases containing low frequency nouns were more easily recognised than phrases containing high frequency nouns which suggests that the recognition of phrases relies on experience of both the entire sequence and of the component words.

Finally, there is electrophysiological evidence of a phrase frequency effect. Siyanova-Chanturia, Conklin, Caffarra, Kaan & van Heuven (2017) looked at ERPs elicited by English highly frequent and predicyable binomials (e.g., *knife and fork*) compared to the infrequent equivalent (e.g., *spoon and fork*) and semantic violation equivalent (e.g., *theme and fork*). It was found that binomials elicited larger P300s and smaller N400s compared to the other conditions. The authors argued that frequent multi-word expressions experience pre-activation of their mental template resulting in a reduced processing load. Furthermore, when speakers were presented with the same stimuli without '*and'*, no activation differences were observed between binomials and the infrequent and semantic equivalent, which provides further evidence for frequent phrases being stored as a single unit.

The above evidence shows that there is reduced processing for high frequency phrases during tasks which rely on comprehension. It seems plausible that this reduced processing is a result of chunking, where sequences which are commonly accessed are stored as a chunk to compensate for the limited resource of working memory. Under such an account frequent sequences are accessed more efficiently and thus have a reduced processing load. However, it is less clear whether any such chunking results in holistic processing or whether individual words are also accessed (see Jacobs et al, 2016, Sosa & MacFarlane, 2002).

1.4.2 Production tasks

Evidence for advanced processing of phrases which are high in frequency have also been well established in production tasks for both adult and child populations. One method which has often been implemented to explore the effect of phrase frequency on production is elicited imitation.

Bannard and Matthews (2008) tested 2- and 3-year-old children's knowledge of frequent multiword sequences via a repetition task using pairs of sequences that were identical except for the final word. Frequently occurring chunks in language (e.g., *a piece of cheese*) were matched to infrequent sequences (e.g*., a piece of food*), where the final words (e.g., cheese, food) and final bigrams (*of cheese, of food*) were matched for frequency. Here, children were found to repeat the first 3 words of the high-frequency word sequence combinations more quickly and more accurately than the low-frequency combinations. This finding has since been replicated using different stimuli with older children. Kueser and Leonard (2020) also used 4-word sequences, which were manipulated for phrase frequency and predictability. Like Bannard and Matthews, the authors found that both typically developing children and children with developmental language disorder made fewer errors when producing the first 3 words of high frequency sequences in comparison to low frequency sequences.

One interpretation of these findings is that frequent use of phrases leads to chunking and that chunked multi-word sequences have a processing advantage. This is consistent with the usage-based model (e.g., Bybee, 2006). This model contends that there is no distinction between the lexicon and grammar. Instead, production is driven by statistical learning. Therefore, inflected words (e.g., *walked*) which are used frequently are stored in the same way as their uninflected counterparts (e.g., *walk*). According to this model, the strength of the representation of a given chunk is determined by its frequency, so that more frequent chunks have more robust representations. Words can also fuse into larger units if they co-occur together often enough. These units can then be accessed directly instead of having to be formed. This model contends that the words within these multi-word sequences can still be accessed, so the storing of these multiword units is not holistic.

The above effect has been replicated in adults. Using a similar methodology to Bannard and Matthews (2008), Arnon and Cohen Priva (2013) asked adults to repeat high (*don't have to worry*) vs low (*don't have to wait*) frequency phrases of the same syntactic type, where word, bigram and trigram frequency were held constant. Adults showed reduced phonetic duration for high frequency phrases compared to low frequency phrases. Furthermore, this effect held for multi-word items in spontaneous speech, regardless of constituency status. Later, Arnon and Cohen Priva (2014) replicated this finding using trigrams. The authors further showed that this phrase frequency effect was not driven by the final word but the frequency of the whole trigram and that when controlling for predictability of the former and preceding words in the sequence, there was still an effect of phrase frequency. In addition to this, it was found that the effect of word frequency was reduced (but did not disappear) for high frequency phrases.

The authors argue this is evidence of phrase and word frequency influencing production simultaneously and these findings are consistent with the usage-based model, which contends that frequent sequences are represented together as units, though the word information is still available. Therefore, speakers have frequency-dependent knowledge of both phrases and words, which might be contained in complementary representations.

Picture naming tasks have also been used to measure fluency in adults for phrases manipulated for frequency. Janssen and Barber (2012) presented native Spanish speakers with line drawings of adjective-noun phrases (colour-object pairs e.g., 'zapato rojo') and noun-noun phrases (e.g., 'zapato rodillo') that varied in phrase frequency (while the object name's frequency was held constant e.g., 'zapato') and measured speech onset latencies. It was found that Spanish speakers were better at producing more frequent adjective-noun and noun-noun phrases compared to their less frequent counterparts. The authors further argued that since the average phrase frequency of noun-adjective phrases tends to be much higher than noun-noun phrases, and given that the observed phrase frequency effect was found in both phrase types then the phrase frequency effect is found ‘across the entire frequency continuum’.

In a second experiment, Janssen and Barber presented French speakers with sets of coloured objects and were asked to produce adjective-noun and determiner-adjective-noun phrases. If the semantic integration of the object and colour influences production, results for determiner-adjective-noun would be the same as adjective-noun phrases even though the phrases have different frequencies. However, the more frequent adjective-noun and determiner-adjective-noun phrases were produced more quickly than the less frequent phrases which suggests the frequency effect is not due to semantic integration of the object and colour. Furthermore, transitional probabilities did not have an effect. Finally, the frequency of the object name did not have an effect on speed of production in either experiment. The authors take these results as evidence that phrases are stored holistically, where there is no interplay between phrasal representations and the phrase and the component words.

The Janssen and Barber study was the model for a study by Shao, van Paridon & Poletiek, (2019) who looked at the production of Dutch adjective-noun and determiner-adjective-noun phrases. Here, utterance onset latencies were shorter for high frequency phrases in both conditions. However, unlike in Janssen and Barber (2012), latencies were also affected by the frequencies of the adjectives in adjective-noun phrases and of the nouns in determiner-adjective-noun phrases, which supports the view that speakers do have frequency derived knowledge both of phrases and of their component words. These findings suggest that speakers might have frequency-dependent knowledge of both phrases and words, which are contained in complementary representations.

Zimmerer, Newman, Thomson, Coleman and Varley (2018) provide evidence for phrase frequency effects in patients with aphasia. Using semi-structured interviews to encourage naturalistic speech, patients with aphasia were found to use more frequent and strongly collocated phrases than non-aphasic speakers. The authors therefore argue that phrases may be stored as holistic units which makes them resilient to aphasia. This further supports the idea that formulaic language (non-literal expressions that are often fixed in form e.g., idioms) is neurologically distinct from novel speech. For example, Van Lancker Sidtis (2012) suggests that the left hemisphere modulates novel speech whereas formulaic language depends on the right hemisphere.

Jacobs, Dell and Bannard (2017) looked at the effect of phrase frequency on recall memory, which is taken to be a production task given that speakers must retrieve and produce stimuli. Speakers were shown adjective-noun phrases which had been manipulated by phrase frequency and were asked to write down as many phrases as they could remember. Frequency did not predict recall of at least one word, but speakers were more likely to recall both words from high frequency phrases given recall of at least one word. This suggests that once one word is retrieved from the phrase, the phrase is more likely to be completed in its entirety for high frequency items. The authors therefore suggest that speakers store phrasal representations consisting of constituent words, held together by an associative connection.

Finally, there is also electrophysiological evidence for phrase frequency effects in production. In a free recall task, Tremblay and Baayen (2010) showed that the probability of the whole phrase, and the frequency of the trigram and constituent words predicted recall. Additionally, the authors claim that ERPs seen during recall show that 4-word sequences were retrieved as a chunk. In previous research, the earliest probability/frequency effect on ERPs for words were observed 110–180 msec after stimulus onset. In this experiment whole-string probability modulated P1 and N1 amplitudes 110–150 msec after stimulus onset. Therefore, it is unlikely that word forms could be retrieved from the lexicon and then be constructed into 4-word sequence within this time frame. The authors took these findings as evidence for 4-word sequences being stored as wholes and as parts.

Hendrix, Baayen and Bolger (2017) used a primed picture naming task to look at the effect of word and phrase frequency on the production of prepositional phrases using ERPs. Speakers were presented with preposition and definite article primes (e.g., 'on the') followed by the target picture (e.g., 'strawberry'). Different ERPs were elicited for words and phrases. For word frequency, oscillations near the lower edge of the theta range across the left hemisphere as well as in bilateral occipital-parietal areas were observed; these areas are thought to reflect working memory demands in language tasks. For phrase frequency, there was an effect over time where more positive voltages were elicited for low frequency phrases, and more negative voltages were elicited for high frequency phrases in the left lateralised parietal and occipital areas. The authors argue that word representations cannot be stored and accessed as each other given the different observed ERPs.

Although there is debate as to the way in which phrases are stored and retrieved, the above findings show that phrase frequency effects do exist for many different experimental paradigms. The implication of the phrase frequency effects for models of production will be considered in section 1.6.

1.5 Phonological complexity

The term 'phonological complexity' refers to various measures used in the literature which aim to quantify how complex a sound/syllable or word, or in some cases whole languages, might be to understand or produce. We are interested in how complexity affects production in children and adults. The effect of phonological complexity on production is relevant because although less complex words should be easier for children to learn to produce, speakers must learn to produce language in a way which is most useful for them to communicate. Schwartz and Leonard (1982) show that 1-year-old children are selective in their production of words based on phonological properties. Here, children were found to produce new words constructed by the researcher more often and more quickly when they contained consonants that they used, in comparison to new words which contained consonants that they did not use. However, speakers do not necessarily learn to produce the words which are most simple most quickly. The following aims to explore different measures of complexity which have been used in the literature, as well as their effect on production.

1.5.1 Comparing the phonological complexity of different languages

Much of the early research on phonological complexity focussed on comparing the overall complexity of different languages, mostly by means of a single measure (i.e., syllable count: Maddieson, 1984). While this is not directly related to the topic of this thesis, many of the ideas here are like those used in developmentally focused measures. Segmental phonemes are a popular focus, with emphasis placed on consonants – given that consonants vary in articulatory complexity and tend to be more difficult to produce than vowels.

Maddieson (2006) established a measure of syllable complexity based on the number of consonants contained in a syllable. This measure was used to compare the complexity of different languages – languages which only allow for CV syllables are thought to be simple (e.g., Maori). Languages which allow for onset consonants, onset clusters and coda consonants are thought to have a moderate level of complexity (e.g., Somali). Languages considered to be complex allow for the above as well as coda clusters (e.g., French).

Table 1-1 Syllable complexity by consonants (Maddieson, 2006)

	Onset Consonant	Onset Cluster	Coda Consonant	Coda Cluster
Simple	✓			
Moderate	✓	✓	✓	
Complex	✓	✓	✓	✓

Many early measures of phonological complexity also place emphasis on consonants and consonant clusters, where words containing more consonants/consonant clusters are thought to be more complex than words containing few or no consonants/consonant clusters.

1.5.2 Measures of phonological complexity in the study of development

1.5.2.1 Form-based measures of complexity

One straightforward metric that has been used to measure complexity is length - utterances which are longer are likely to contain more sounds/syllables and therefore be more difficult to produce. Brown (1973) used Mean Length Utterance (MLU) as a measure of language proficiency. Here, Brown suggests taking 100 spoken utterances from a child's spontaneous speech and dividing the total number of morphemes produced by 100 to give the average number of morphemes per utterance. This then gives a score which can be used to gauge the complexity of a child's speech.

This was later expanded by Ingram (2002) who used the Phonological Mean Length of Utterance (PMLU) as a measure of complexity. The PMLU measures the length of a word and, like the above measures, places emphasis on the number of consonants correctly produced.

Each word receives points based on:

- The number of segments produced by the child per word (each consonant and vowel receive one point)
- The number of consonants produced correctly (each correct consonant receives one point)

The point total for each word is the sum of the two measures. Also calculated is the proportion of the whole word correctly pronounced (Proportion of Whole-Word Proximity), which is calculated by dividing the child's PMLU score by the maximum possible PMLU score of the target word. This score can also be used to measure the complexity of a child's language. However, there is no measure for when a child adds segments to the target word and the correctness of the vowel is also ignored.

These early ideas were later integrated into a more detailed measure of complexity, the first of which was developed by Jakielski (1998). The Index of Phonological Complexity (IPC) is an unpublished (other than a conference poster) tool based on the phonetic analyses of infant babbling for 5 observed infants (aged 7-36 months). In this measure, a value is assigned to productions based on 8 different factors, so that productions with higher scores are thought to be more complex (see Table 1-2). For example, words containing fricatives, affricatives and liquids are thought to be more complex than words containing stops, nasals, and glides, while words ending with a consonant are thought to be more complex than words ending with a vowel, and words that are longer (e.g., containing at least 3 syllables) are thought to be more complex than words made up of a monosyllable or disyllable.

Table 1-2 Index of Phonetic Complexity

Factor	No Score	One point awarded each
Consonant by Place	Labials, coronals, glottals	Dorsals
Consonant by Manner	Stops, nasals, glides	Fricatives, affricatives, liquids
Singleton Consonant by Place	Reduplicated e.g., VC-VC	Variegated e.g., VC-CV
Vowel by Class	Monophthongs, disyllables	Rhotics
Word Shape	Ends with a vowel	Ends with consonant
Word Length	Monosyllables, disyllables	At least 3 syllables
Contiguous Consonants	No Clusters	Consonant Clusters
Cluster by Place	Homorganic	Heterorganic

Stoel-Gammon (2010) later developed the Word Complexity Measure (WCM) focusing on early phonological acquisition in English (using speech samples from 7 children aged 7-48 months). Productions which require later acquired phonological parameters score more highly on the WCM and are thought to be more complex. For example, sound classes like stops, nasals and glides are normally established earlier in language compared to sound classes like fricatives and affricates which are acquired later in development. Therefore, points are awarded to fricatives, affricates, and other later acquired (and therefore more complex) sound classes (see Table 1-3).

Table 1-3 Word Complexity Measure

Factor	One point awarded
Word patterns	More than 2 syllables
	Stress on any syllable but the first
Syllable structures	Word-final consonant
	Consonant cluster (sequence of two or more consonants within a syllable
Sound classes	Velar consonant
	Liquid, syllabic liquid, rhotic vowel
	Fricative, Affricative
	Voiced fricative or affricative

*One point given for each sound class in production

Using this measure, the target word ‘baby’ (beɪbi) would be awarded a score of 0. In the case of word patterns, this production is less than 2 syllables, and has stress on the first and only syllable. For syllable structures, the target does not have a word-final consonant, or consonant cluster, so is not awarded any points. Finally, ‘baby’ does not contain sound classes which are given a point. Therefore, by this measure, the word ‘baby’ is not thought to be a particularly complex production. In comparison, the target word ‘chicken’ (tʃɪkɪn) would be awarded 3 points. This production requires a word-final consonant and contains both an affricate and velar sound class. Therefore, the word ‘chicken’ is thought to be more complex than the word ‘baby,’ and you might expect children to find it more difficult to produce the word ‘chicken’ in comparison to the word ‘baby’.

1.5.3 Input-based accounts of phonological complexity:

The above measures quantify complexity using factors like sound class and syllable structure. This is reasonable given that some sounds are intrinsically harder for the phonological system to articulate. For example, producing a fricative (e.g., ‘f’) requires more complex movement of the vocal system than producing a stop (e.g., ‘b’). It is also assumed that sounds which are intrinsically difficult to produce are produced later in development. Measures like the IPC and WCM use this to calculate complexity. However, it is also possible that complex sounds occur later in development because they are not used as often. A consonant which is rare in language might have lower accuracy rates as it has not been practiced enough. There would also be less opportunity to gain abstract knowledge of the consonant away from known words, which makes the learning of new words more difficult. Therefore, when determining the complexity of a given sound or word, it is also important to consider its usage in language.

Maddieson (2009) suggests that the complexity of a language varies according to the frequency of more complex components and proposed a measure for this (multiplying the rank frequency of phoneme by its level of complexity (Lindblom & Maddieson, 1988)) to give a comparable score. In this case, rank is expressed as decreasing decimal fractions of 1 (i.e., 1, 0.9, 0.8, 0.7, etc). Maddieson uses the example of the segment /n/ in English which would contribute 0.8*1 i.e., 0.8 to the score. While the segment /z/ would add 0.3*2, i.e., 0.6.

1.5.4 Research evidence for effects of phonological complexity effects

1.5.4.1 Phonotactic probability

Another measure which considers complexity in terms of frequency is phonotactic probability. Phonotactic probability refers to the likelihood of a sound occurring in a specific utterance position. There are different measures of phonotactic probability. For example, segment positional probabilities refer to the likelihood of phonemes occurring in the onset, vowel and coda position (CVC), while biphone probabilities refer to the likelihood that phonemes are preceded or followed by another specific phoneme. Phonotactic probability can also refer to the probability of a phoneme occurring in a certain position within a word.

Since speakers are sensitive to input frequency, words with low phonotactic probability are likely to be more difficult to say and therefore could be argued to be more complex. In fact, it has been shown that both adults and children do tend to find words with low phonotactic probabilities more difficult to produce and this preference is observed from a very early age. By 7 to 9 months, infants show a preference for CVC nonwords with high phonotactic probabilities in head turning procedures (Zaumuner, 2003; Jusczyk & Charles-Luce, 1994).

Zaumuner, Gerken and Hammond (2004) measured 2-year old's repetition of coda consonants contained in CVC non-words which were produced during a picture naming task using imaginary animals. Stimuli contained identical codas across the 2 conditions while neighbourhood densities and word-likelihood ratings were controlled for. Here, infants were more likely to produce the coda in the high phonotactic probability condition. Furthermore, 3- to 6-year-olds are better at producing nonwords with high phonotactic probability compared low phonotactic probability (Storkel, 2001) and adults were found to take less time to produce CVC non-words containing high phonotactic probability patterns in an elicited imitation study (Vitevich and Luce, 1998).

Although some authors have argued that this phonotactic probability effect is simply a consequence of phonetic complexity and phonotactic probability being highly correlated, findings from Goldrick and Larson (2008) suggest that this is not the case. Here, adults were presented with an implicit learning paradigm containing novel phonotactic constraints. In this experiment phonetic complexity and phonotactic probability were uncorrelated and phonotactic probability was shown to still influence production independently of phonetic complexity.

1.5.4.2 Neighbourhood density

Another measure which takes frequency into consideration is called phonological neighbourhood density (PND). PND is calculated by the number of words in the language that have phonetically similar words (e.g., rat, bat, etc.). Since speakers are sensitive to frequency, words with low neighbourhood density are likely to have been produced less often, which means that speakers might find them more difficult to say, which in turn makes them more complex. In two speech-error elicitation tasks (Vitevich, 2002), adult speakers were found to make more errors for words with sparse neighbourhoods compared to words with dense neighbourhoods and in a further three picture naming tasks, words with sparse neighbourhoods were produced more slowly than words with dense neighbourhoods. Munson and Solomon (2004) also found that adults produced words with shorter durations and more expanded vowel spaces when they were high in PND, while children aged between 0;11 and 4;0 are more accurate and less variable in spontaneous productions for words which are high in PND and frequency (Jones, 2020).

1.5.4.3 Stuttering and phonological complexity

Some authors have suggested that phonological complexity predicts stuttering. Anderson (2007) used word frequency, PND and neighbourhood frequency as measures of complexity to explore the effect of complexity on stuttering during naturalistic speech in preschool children. Neighbourhood frequency refers to the average word frequency per 1 million of all the word’s phonological neighbours, divided by the number of neighbours. It was found that children stuttered less often when words were high in word frequency and neighbourhood frequency, but PND was not found to affect stuttering rates, though the author suggests that this was due to more of the stuttered words being function words.

More recent measures of phonetic complexity such as IPC and WCM have most often been used to measure fluency in speakers who stutter. For example, IPC scores predicted stuttering rate in both German (Dworzynski & Howell, 2004) and Spanish (Howell & Au-Yeung, 2007) speakers for content but not function words. LaSalle and Wolk (2011) also found that dysfluent words had higher IPC scores and fewer phonological neighbours compared to matched fluent words in the spontaneous speech of 3 14-year olds diagnosed as a stutterer, a clutterer and a stutterer-clutterer. Furthermore, Wolk and LaSalle (2015) found that in combined online speech samples from speakers aged between 12 and 19 with stuttering or with stuttering and concomitant language disorder, these speakers were less fluent for words which were marked to be higher in the IPC and WCM and also lower in density, compared to fluent words.

1.5.5 Phonological complexity and word frequency

We know that frequency has an effect at the word level (see section 1.3) where frequent words are produced more fluently than infrequent words. However, experiments which show that speakers are more fluent in their production of words which are high in phonotactic probability and neighbourhood density suggest that frequency also plays a role in how complex speakers find different sounds and words, where practiced sequences are produced more fluently. Sosa and Stoel-Gammon (2012) provide evidence for this idea. During an elicited imitation task 2-year-old children were found to be more variable in their production of more complex words (containing later developing consonants and syllable structures). But complex words which were high in frequency were found to be less variable than complex words which were low in frequency. Sasisekaran et al (2010) found that both adults and children showed a learning effect (see 1.3.2) when presented with non-words of varying complexity (complexity was measured by syllable number, age of acquisition for included consonant/consonant clusters and length of consonant clusters), so that the production of complex non-words became more fluent after practice.

One explanation for this finding, where speakers become more proficient in their production for complex words that have been practiced more often, is called phonological/motor chunking (discussed in more detail during section 1.6). The idea is that frequent sounds are represented together as a single unit. This means that the production system requires a reduced working memory load to access the target form. In addition to this, chunking is said to also occur at the motor level – often called a 'motor program.' Here, sequences are produced more fluently because the articulatory program has been practiced more often in comparison to less frequent sequences.

One study which provides direct evidence for phonological/motoric chunking was conducted by Segawa et al. (2015), who explored the trade-off between phonological complexity and frequency using phonotactic probability as a measure of complexity. Adults were presented with phototactically legal and illegal triconsonantal initial and final consonant clusters contained in mono-syllabic non-words. It was found that speakers had a significantly lower error rate after practice for the illegal sequences but did not differ in their repetition of legal syllables. The authors took this latter finding as evidence for learned motor programs - given that performance improvement was specific to the stimuli encountered, production could not be a result of the learning of phonological rules, instead improvement is a consequence of motor learning.

In conclusion, it has been shown that both adults and children are more successful in their production of words which are less complex in comparison to words which are more complex. However, speakers are more fluent in their production of complex words which are more frequent compared to complex words which are less frequent. To explore this effect further, models of production which consider phonological/motor chunking will be discussed in section 1.6.

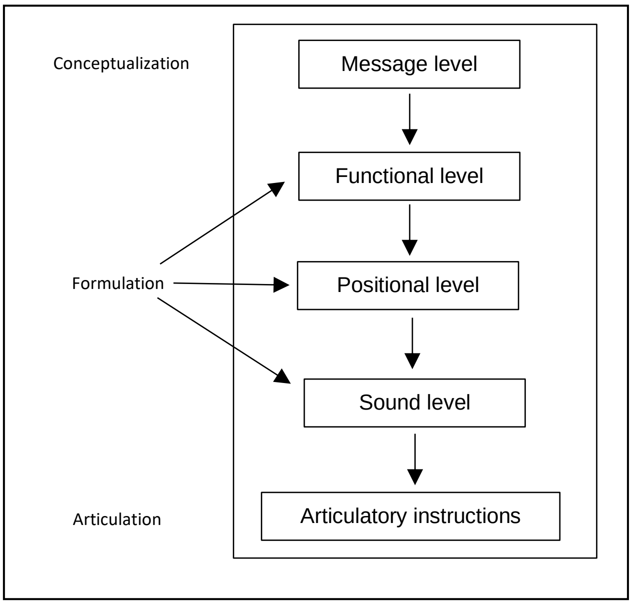
1.6 Models of production

Models of production consider the processes involved in the retrieval and subsequent production of a target word or sequence. Although models of production have mostly focused on the production of syllables/words, the following will discuss these models in terms of both words and phrases. Emphasis will also be placed on how the processes underlying these models drive frequency effects, specifically focussing on word and phrase frequency.

1.6.1 Garrett's (1975) model of production

The earliest widely cited model of production was developed by Garrett (1975). This model was based on the analysis of speech errors.

Figure 1-1 Diagram of Garrett's (1975) Model of Production



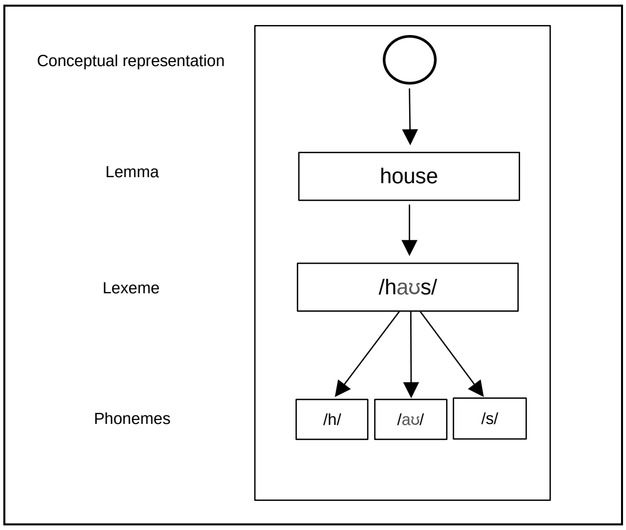
In this model, the conceptualisation of the message is realised at the message level. Here, the speaker retrieves the 'idea' of the target word that they would like to produce. Next, the message level converts this into the linguistic form. During the functional level, semantic information and content words are accessed. Then, at the positional level, the sentence structure is planned, function words are retrieved, and words are placed in relevant positions. Next, the phonological form of the message is retrieved at the sound level before the final stage of articulation. This processing is serial with no interaction between the different stages.

Garrett argued that word exchange errors occur because words are retrieved before their position has been defined. Furthermore, content and function words are retrieved separately since errors only occur in the same syntactic class. While exchange errors occur because sounds are not specified until later. This model was the first to fully consider the production process and many of these ideas were maintained in future models. This model does not account for word or phrase frequency effects, but shaped many future models that attempt to do so.

1.6.2 Two-stage model of lexical access (Levelt, 1989; 1999)

Levelt (1989; 1999) proposed a two-stage model of lexical access which, like Garrett, assumes modularity. This means that each stage is distinct – the information accessed during each stage is restricted to that stage. This model (depicted in Figure 1-2) uses top-down processing and is not interactive. Therefore, there is no activation of the second stage until activation of the first stage has been completed. Once the target word has been produced the activation ends. In this model, the speaker first specifies a concept, which leads to activation of the linguistic form (lemma). The lemma also represents a word's syntactic and semantic information. Here, other lemmas with similar contexts are also activated. There is competition between syntactically and semantically related lemmas while lemma selection occurs. Once the target lemma has been selected, the lexeme (phonological form of the word) is realised before the final stage of articulation, which involves motor execution and production of the word (see Figure 1-2).

Figure 1-2 Diagram of Levelt's (1989) Model of Production



Each stage of this model is associated with a different type of error. Semantic errors are thought to occur during lemma access in the formulation stage (e.g., saying ‘fingers’ instead of ‘toes’), while phonological errors occur during the articulation stage (e.g., saying ‘gat’ instead of ‘cat’). Therefore, mixed errors (an error which includes both semantic and phonological e.g., saying start instead of stop) would only occur by chance since semantic and phonological units are modular and separate.

Word frequency effects in this model are attributed to the second (phonological) stage of lexical access. The speed of access to phonological forms is constrained by word frequency, so that speakers are faster to access the phonological forms of words which are more frequent, compared to the phonological forms of words which are less frequent (see Levelt, 2001 for description).

Evidence for this was taken from a series of studies conducted by Jescheniak and Levelt (1994). In one of the experiments reported (experiment 6), the authors used homophones to show that the speed of production for a given word is affected by the frequency of its homophone pair (homophones have an identical phonological form e.g., ‘genes’ and ‘jeans’). In this experiment, speakers produced homophones at the same speed as control words matched for the combined frequency of both homophones, which was faster than their production of control words matched for independent word frequency. The authors took this to mean that low-frequency homophones inherit the fast access speed of their high frequency partner and therefore the word frequency effect is held to occur at the formulation stage.

In experiments 4 and 5, participants were asked to perform a gender decision task for line drawings of simple objects manipulated for frequency. Speakers were asked to decide on the grammatical gender that the depicted object's name takes and push a ‘de’ or ‘het’ button according to their decision (in Dutch masculine and feminine words are *de* words, whereas neuter words are *het* words). The authors did not predict an effect of word frequency on participant response latencies given that the task requires retrieval of the lemma but does not require retrieval of the noun's phonological form, since the noun is not produced. Their results were in line with this prediction.

However, there is some evidence to suggest that word frequency does have an effect at the semantic level. Navarette, Basagni, Alorio and Costa (2006) presented Spanish speakers with pictures of objects (e.g., a picture of a car) and asked them to make a button press decision on the grammatical gender of the names of pictures. In this case adult's responses were faster for pictures with high-frequency names compared to low-frequency names. By the same logic as the above experiment, the authors argue that this provides evidence for a frequency effect at the lemma level.

The tip of the tongue (TOT) phenomenon has also been used as support for a two-stage model of lexical access. It is argued that TOTs occur when the first stage of lexicalisation is activated, but the second stage is not. Since semantic and phonological processing are distinct processes, the speaker is caught between stages and cannot produce the word, while feeling that they know information about the word. Furthermore, TOTs are more likely to occur in low frequency words, with fewer phonological neighbours (Harley and Brown, 1998).

1.6.3 Levelt's two-stage model and phrase frequency

Although Levelt and Meyer (2000) do briefly consider how the 2-stage theory might allow for the retrieval of multi-word sequences, this model of lexical access does not explicitly account for phrase frequency effects. In this paper, the authors suggest that in order to produce a phrase, nouns are retrieved and selected first, before the syllabification and phonological encoding for each word, which ensues incrementally. However, the authors do not state whether articulation begins after the first word has been planned or whether articulation occurs after the entire sequence has been planned.

Evidence for this was taken from a study by Schriefers (1993) who found that Dutch speakers cannot complete form encoding (retrieval of a words phonological information/articulatory gestures) for an adjective without having accessed the lemma of the noun. This is because the gender of the noun must be marked on the adjective. In a picture/word dependency task, subjects were required to name simple Dutch phrases (e.g., red chair) while ignoring an auditory distractor noun which was gender congruent or gender incongruent. When presented with a gender incongruent distractor noun, naming latencies for the entire phrase were found to be longer.

The authors also consider how their model might allow for sentences containing 2 or more nouns. They propose that the retrieval stages could be modular or interactive. Therefore, initiating the articulation of the first word could be dependent on accessing the second word. However, they also suggest that the processes could run in parallel. Therefore, there are no specific predictions for how phrase frequency affects production, emphasis is instead placed on word frequency.

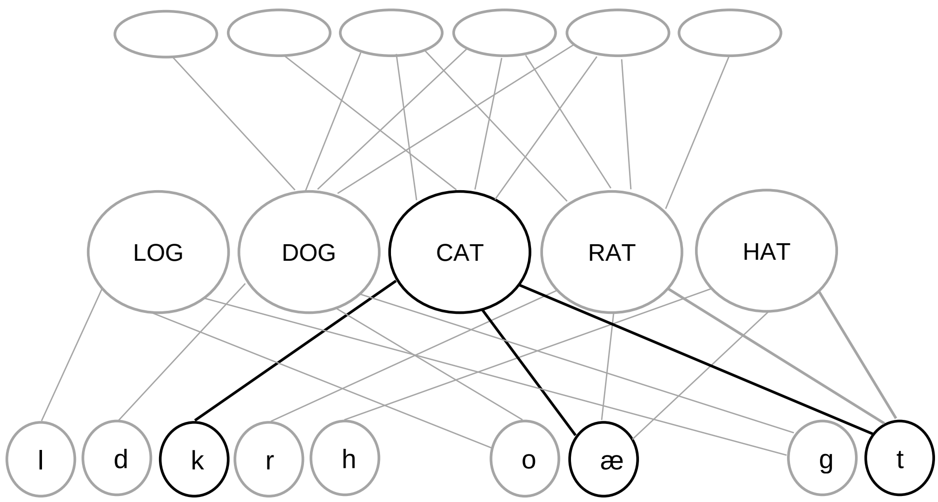
1.6.4 Interactive two-step model (Dell, Schwartz, Martin, Saffran & Gagnon, 1997)

The interactive two-step model consists of two stages, one of which deals with word retrieval and the other with phonological retrieval (see Figure 1-3). This model differs from that of Levelt in that it allows for the bidirectional spread of activation between the model's two steps. Therefore, phonological information can influence word retrieval, and vice versa. Selection is based on the weight or strength of connections and decay is the only means by which activation levels are reduced.

The word retrieval stage is concerned with initial selection of the target word. To retrieve and produce a target word (e.g., CAT), first the semantic units of the target are given a jolt of activation which spreads through the network. The target word 'CAT' is activated along with semantically related words (e.g., 'DOG' and 'RAT'). Phonologically related words (e.g., 'HAT' or 'RAT') are also activated from the bottom-up spread of activation due to shared phonemes. The word with the strongest activation is then selected for production. In most cases the target word will have the strongest activation, but it is possible for a semantically related word (e.g., DOG), mixed word (e.g., RAT), phonologically related word (e.g., HAT) or even an unrelated word, to be selected instead of the target. Since selection is based on the anticipated syntactic structure of the sentence, the system will always select a word of the same syntactic category.

Next, during the phonological retrieval step, the target word (e.g., CAT) is activated while other nodes retain their residual activation from the previous step. This means that activation for the target word is stronger than that for the other related words. Once 'CAT' is activated, further activation spreads through the network in both directions, towards phonemes and back up towards semantic units, before the most active phonemes are selected. Errors during this step are most likely to lead to phonologically related words (e.g., MAT) or non-words (e.g., CAG). Finally, after phonemes have been selected, articulation takes place.

Figure 1-3 Interactive 2 stage model (Dell et al, 1997)



This model contends that many of the errors which occur in production are a result of the unconstrained activation that spreads bidirectionally. Since multiple nodes can be activated at once, speech errors occur when an incorrect item has a higher level of activation than the target item. For example, exchange errors (e.g., ‘writing a mother to my letter’ vs ‘writing a letter to my mother’) occur when the syntactic frame has been constructed, but ‘mother’ is activated before ‘letter’. After selection, the activation level for ‘mother’ returns to 0 and the activated node ‘letter’ is selected for the second noun slot.

Similarly, blend errors (e.g., letter/note -> lote) occur when both nodes are equally active, so both nodes are selected which affects the lower encoding levels, resulting in a blend of the word’s sounds. Feedback from the phonological stage to the word retrieval stage can result in the lexical bias effect, where the target word is replaced with another phonologically similar real word (e.g., kitten vs mitten). Finally, the prevalence of mixed errors (an error that contains both semantic and phonological e.g., start vs stop) supports cascading models, like this interactive model, and contradicts discrete models (e.g., Levelt, 1989), given that these errors occur more often than the modular models predict in both naturalistic speech (Dell & Reich, 1981) and picture naming tasks (Rapp & Goldrick, 2000).

Levelt, Schriefers, Vorberg, Meyer, Pechmann and Havinga (1991) dispute that production processes are interactive. The authors agree that there is competition between semantically related words during lemma selection and found that semantic activation primed phonological activation for target items (e.g., ‘sheep’ primed ‘goat’). However, Levelt et al. argue against bottom-up spreading activation. It was found that phonological activation did not prime other semantically activated items as predicted in activation spreading models. (e.g., ‘sheep’ did not prime ‘goal’ via activation of ‘goat’). However, Dell and Seaghdha (1992) argue that this evidence is compatible with an interactive account since competing words are not normally activated as much as the target word. The model allows for mixed error effects but does not contend that they will occur more often than correct activation of the target word. In addition to this, Peterson and Savoy (1998) found that ‘soda’ was activated during retrieval of the target ‘couch’ (via ‘sofa’).

Although there is disagreement between the models described as to whether the 2 stages of production are independent or interactive, it is generally accepted that there is a semantically/syntactically specified representation followed by a phonologically specified representation. This system is organised by frequency, with more frequent words/sounds being more easily accessed. In the above interactive model, since activation is based on the weight or strength of connections, words which are more frequent are thought have stronger weights. Therefore, the frequency effect is located at the level of lexical selection (e.g., Dell, 1990). Based on the assumptions in this model, it is also reasonable to expect that more complex words (which contain more phonemes) will be more likely to result in an error (where one of the phonemes will fail to be accessed during the phonological retrieval step). As for phrase frequency, this model follows the words and rules assumption (e.g., Pinker & Ullman, 2002). Therefore, this model does not make any predictions for phrase frequency.

1.7 Neural network models of production

1.7.1 GODIVA model (Guenther, 2016)

The Gradient Order Directions into Velocities of Articulators (GODIVA) model was developed as an extension of the DIVA model (e.g., Guenther, Ghosh, & Tourville, 2006). This model simulates production using a neural network. The GODIVA model can be broken down into 2 interacting parts (see Figure 1-4). Within these parts, multiple items can be active at once. Each item is represented by an activity gradient so that items are selected based on this, with the most active items being selected first.

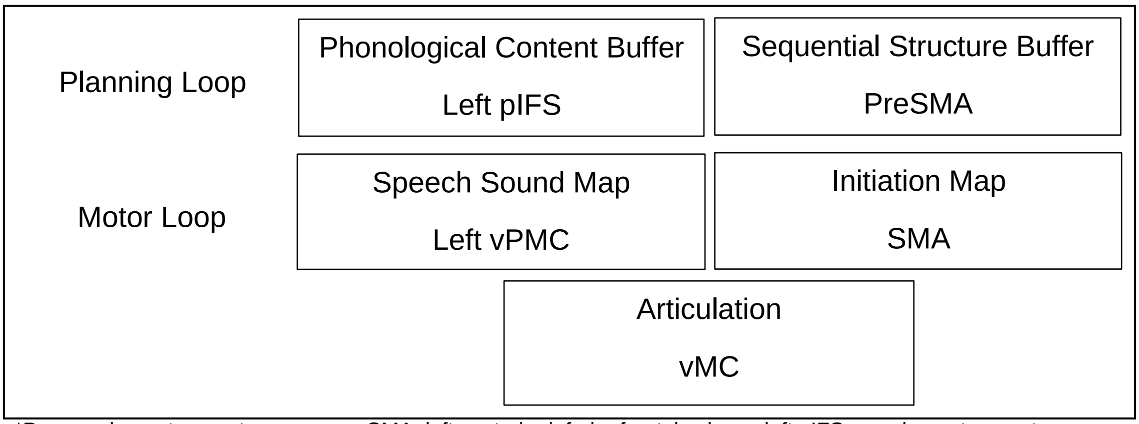
The first part is the Planning Loop, which contains the Sequential Structure Buffer and the Phonological Content Buffer:

- The Sequential Structure Buffer (situated in the Pre-supplementary motor area: preSMA) considers the sequential structure of a target sequence (e.g., word frame structure, syllabic order) and is responsible for the temporary storage of this structure for the target phonological sequence.
- The Phonological Content Buffer (situated in the left posterior inferior frontal sulcus: left pIFS) considers the phonological content of a target sequence (e.g., number of syllables, phonemes) and is responsible for temporary storage of the phonological units of the sequence

The second is the Motor Loop, which contains the Initiation Map and the Speech Sound Map.

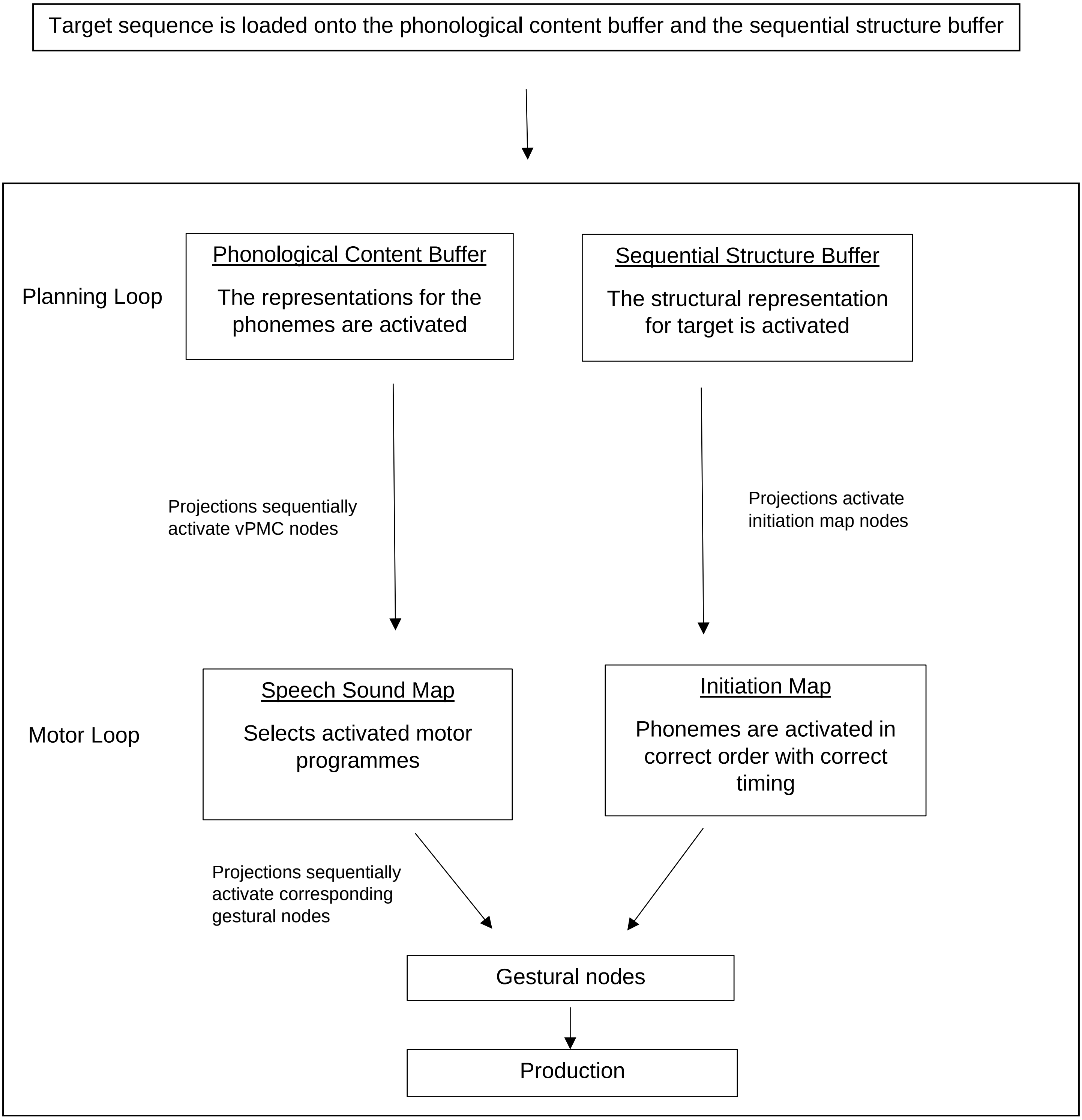
- The Initiation Map (situated in the supplementary motor area: SMA) is responsible for the timing of motor movement/articulation. Here, the order of motor programs is decided by the activity gradient.
- The Speech Sound Map (situated in the left ventral premotor cortex: left vPMC) selects the most active motor program ready to be executed.

Figure 1-4 Simplified diagram of GODIVA model of speech sound sequencing



*Pre-supplementary motor area – preSMA, left posterior inferior frontal sulcus - left pIFS, supplementary motor area – SMA, left ventral premotor cortex - left vPMC, ventral primary motor cortex – vMC.

Figure 1-5 Path to production according to the GODIVA model



The GODIVA model assumes that the production system stores frequently occurring sub sequences as cohesive ‘chunks’ which reduce phonological working memory load and improve motor performance. For non-optimised targets, once the target sequence is loaded, the structure and phonetic content are chosen based on activity gradient. Meanwhile, the sequential structure buffer also projects to the initiation map, which becomes active when the current context for the current sound is recognised. At the same time, the phonological content buffer signals the upcoming phonological items to the speech sound map, which choose the best motor program for producing these items. For non-optimised targets, the speech sound map activates each phoneme in sequence. Finally, articulation occurs when the associated initiation map node is activated.

However, for optimised targets, working memory buffers in phonological content buffer (pIFS) and sequential structure buffer (preSMA) contain cluster sized SSCs to reduce the number of items that are stored in working memory. SSCs refer to the way a syllable is broken into 3 sub syllabic constituents. These are the onset (one or more consecutive consonants at the beginning of the syllable), nucleus (vowel, diphthong or sonorant consonant) and coda (one or more consecutive consonants at the end of the syllable). For optimised targets, the gestures for these SSCs are mediated by the basal ganglia loop instead of the sequential structure buffer (preSMA), which also activates gestural motor programs in the initiation map cells in the initiation map (SMA). Meanwhile, the speech sound map activates the syllable/word as a chunk and subcortical loops in the cerebellum coordinate and coarticulate the individual motor gestures. Finally, articulation occurs when the associated initiation map node is activated.

Segawa, Tourville, Beal and Guenther (2015) provide further evidence for learned motor programs (see section 1.3.2). Here, American English speakers were asked to produce mono-syllabic non-words containing phonotactically legal (*e.g,. BLERK, THRIMF, TRALP*) and illegal (*e.g,. FPESHCH, GVAZF, TPIFF*) bi- or triconsonantal initial and final consonant clusters. Speakers made significantly fewer errors after practice for illegal sequences compared to legal syllables. Speakers also made more errors during repetition of novel illegal syllables compared to these learned illegal syllables. Therefore, performance was not the result of learning phonological rules but of motor learning for the illegal consonant clusters.

This study also provides further support for the GODIVA model as the relevant areas were activated during production. For example, learned illegal syllables resulted in reduced right pre-SMA and left GP activity compared to the production of novel illegal syllables. This is important because both the sequential structure buffer and BG-thalamocortical loop are involved in selecting and initiating phonological chunks. In addition to this, learned illegal syllables resulted in reduced left PMC activity compared to the production of novel illegal syllables (which is due to less speech sound map nodes being activated), while activity in the vMC is approximately the same for learned illegal and novel illegal syllables, because both nonwords require the same articulatory gestures.

Segawa et al. (2019) later expanded these findings. Adults were asked to repeat novel CCVCC non-word phoneme sequences in a practice session, where half of the non-words contained native (phonotactically legal) consonant clusters and half of the syllables contained non-native (phonotactically illegal) consonant clusters. The following day, the speakers were asked to repeat 4 types of CCVCC non-words in a test session: practiced CCVCC, practiced CC, practiced CVC, novel CCVCC. It was found that speakers took less time to produce syllables with non-native consonant clusters after practice, and that this improvement generalised to new syllables that contained those clusters which suggests that improvement was due specifically to the learning of consonant clusters. However, it was also found that speakers made fewer errors for non-words after practicing the whole CCVCC syllable, compared to when speakers had just practiced the consonant clusters. This suggests that learning is also due to practice of the whole syllable. To explain this, the authors suggest that speakers use chunked consonant cluster phonological representations, which are stored in working memory, as well as optimised muscle activation patterns for both consonant clusters and full syllables, which are stored as motor programs.

In this study, the authors make a distinction between phonological (chunked sequences of phonemes) and motoric chunks (a learned motor routine) – both of which are said to occur during the phonological stage of production. In this thesis I will refer to phonological/motor chunking to acknowledge this ambiguity. The experimental effects reported in this thesis are possibly phonological or motoric. Therefore, the term phonological/motor chunking will be used.

This model contends that word frequency has an effect at both the word retrieval and phonological level. During the first stage, speakers take less time to retrieve words containing chunked consonant clusters. During the second phonological stage, there is a processing advantage for whole words which are high in frequency, where speakers make fewer errors for frequent words since they are stored as chunked motor programs. Although this model does not yet consider the effect of phrase frequency, the model also does not explicitly consider words, but motor sequences. Therefore, it could be argued that phrase frequency effects would just be the result of chunks that cross word boundaries.

1.7.2 Dual stream model (Hickock, 2012)

The dual stream model was first developed by Hickok and Poeppel (2004, 2007). This model assumes parallel processing where the ventral stream maps sound to meaning and the dorsal stream maps sound to action. Therefore, the ventral stream is involved in comprehension (recognising/understanding input), whereas the dorsal stream is involved in the processes which underlie production. This model is largely based on findings taken from neuroimaging data for tasks exploring both speech comprehension and production (see Figure 1-6).

During the initial stage of speech processing, when speech is heard by the speaker, first the auditory region of the superior temporal gyrus (STG) is activated bilaterally. Functional imaging evidence show that this includes the dorsal STG and phonological network (superior temporal sulcus – STS). The phonological network is said to be contained in the STS because portions of the STS have regularly been shown to be involved in representing and/or processing phonological information (see Hickok & Poeppel, 2007). For example, there is evidence to suggest that portions of the STS are selective for acoustic signals which contain phonemic information compared to complex non-speech signals.

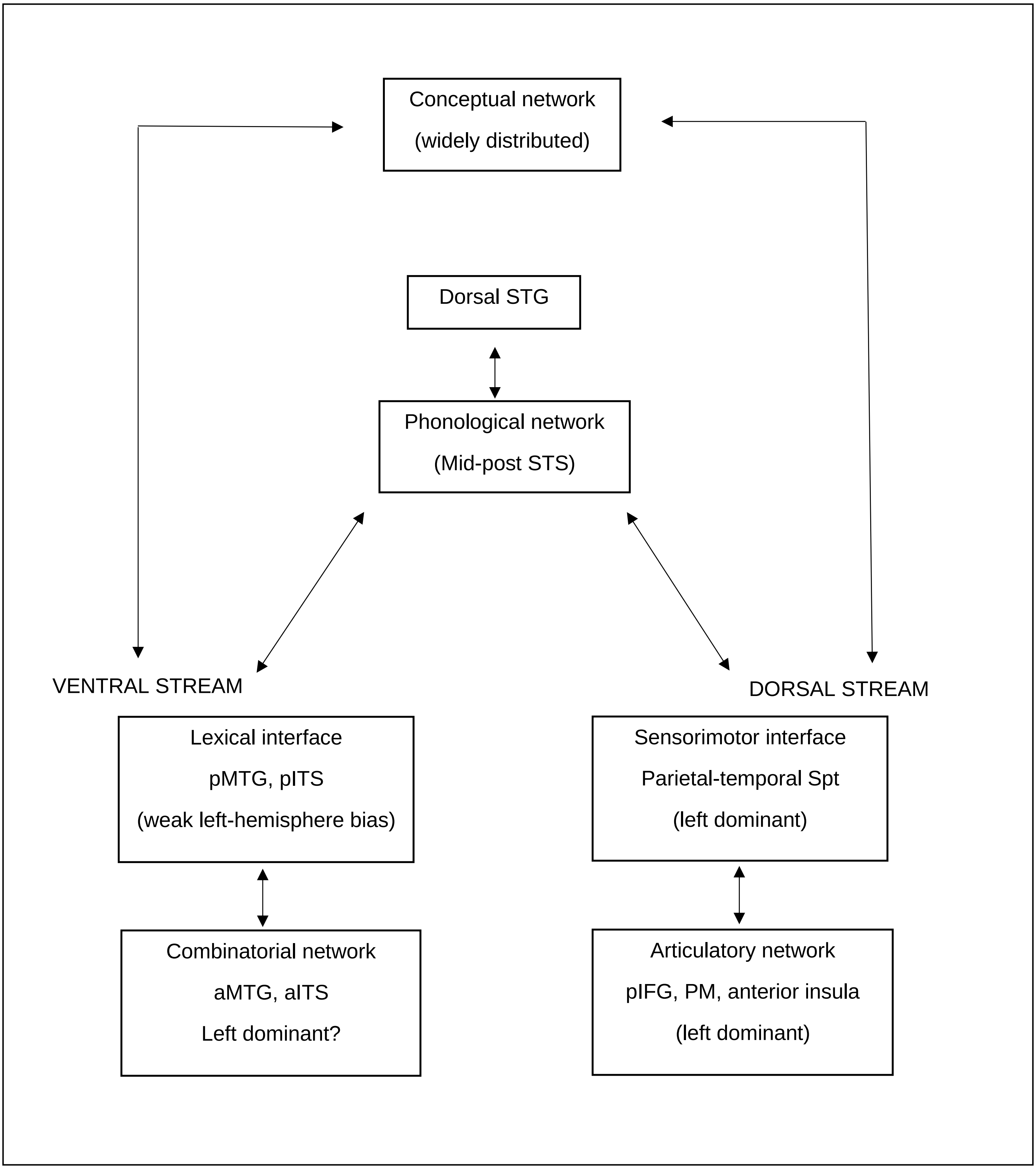
The phonological network next projects to both the ventral and dorsal stream. The ventral stream is a bilaterally organised network which involves portions of the temporal lobe – this stream is said to support speech comprehension processes. Here, the lexical interface uses the auditory input to access conceptual-semantic representations. Although this model proposes that conceptual-semantic representations are widely distributed throughout the cortex, the lexical interface (contained in the temporal lobes) is said to map between this conceptual information, and the phonological information taken from the phonological network. Meanwhile the combinatorial network, contained in the anterior temporal lobe (ATL) regions, mediates lexical-semantic and sentence level processing (syntactic and compositional semantic operations).

On the other hand, the dorsal stream is left-dominant and involves structures in the posterior frontal lobe and posterior dorsal-most aspect of the temporal lobe and parietal operculum. The dorsal stream coordinates links between acoustic speech signals and motor representations. This interaction is important because motor targets are based on auditory targets rather than motor targets (when a speaker produces a word incorrectly, they will change their motor movements to meet the correct auditory target). Here, the phonological network projects to the sensorimotor interface (Spt). The sensorimotor interface is an interactive interface which acts as a sensory-motor integration system. This system can feed forward and backword. Once activated, the sensorimotor interface projects to the articulatory network (contained in the anterior frontal lobe) which coordinates motor representations, before articulation.

Evidence for sensorimotor interface acting as this integration system comes from studies looking at conduction aphasia (where speakers have normal comprehension but experience frequent phonemic errors when speaking). This condition is thought to be a consequence of an interruption of this interface. As phonological representations from the phonological network activate motor representations via the sensorimotor interface, the sensorimotor interface checks the predicted auditory representation against the auditory target. If there is a match, articulation will commence. If there is no match, a correction signal can be generated to activate the correct motor unit. However, speakers with conduction aphasia experience disruption in the sensorimotor interface. Therefore, they cannot generate this forward prediction and so make phonological errors. The lesion does not disrupt the activation of auditory targets via the lexical semantic system which is why patients can detect errors in their own speech. Once this error is detected though, disruption to the sensorimotor interface means that no correction signal can be sent back to change the motor unit and have correct production.

This model is compatible with phonological/motor chunking but also has a clear distinction between semantic processing and realisation, which is compatible with two stage models discussed above. To generate a new word, each step in the path of production is taken. However, the authors argue that once this word is practiced, motor coding becomes automated and production requires little sensory guidance. Therefore, more frequent words are produced more quickly. This model does not consider the production of phrases, or the effects of phrase frequency.

Figure 1-6 Path to production according to the Dual-stream model



1.8 Summary of thesis

The aim of this chapter was to review the existing literatures relevant to three predictors of fluency in speech production - word frequency, phrase frequency and phonological complexity. Both children and adults show a processing advantage for the production of words which are high in frequency. We have discussed some different models of production which aimed to explain this effect. Although there is disagreement as to whether the 2 stages of production are independent or interactive, it is generally accepted that there is a semantically/syntactically specified representation (sometimes called conceptualisation) followed by a phonologically specified representation (sometimes called formulation). In some accounts this system is said to be organised by frequency, with more frequent words/sounds being more easily accessed. There is some debate about at which level this frequency effect occurs. Although most researchers agree that this frequency effect occurs at the phonological level (e.g., Jescheniak & Levelt, 1994), there is some evidence to suggest that word frequency has an effect at the semantic level (e.g., Navarette et al., 2006).

This thesis largely focuses on production frequency, so that any observed effect of frequency is thought to be due to the frequency with which speakers have retrieved/produced the target word/phrase. However, it is possible that input frequency and production frequency are confounded. Therefore, effects related to conceptualization could possibly be related to input frequency (where speakers have more robust conceptual knowledge of words that they have heard more often).

Speakers have also been shown to find complex words more difficult to produce than words that are less complex. However, speakers do improve in their production of complex words when they are also more frequent. One possible explanation for this is some kind of phonological/motor chunking. Although models of production which consider chunking do not necessarily agree on the structure or format of chunks, they generally accept that there is some level of chunking which occurs during the phonological stage of production – possibly at the working memory level – where chunked sequences are treated as a single item in order to reduce processing load and or possibly at the motor level – often called a 'motor program.' Here, sequences are produced more fluently because they have been practiced more often in comparison to less frequent sequences. The GODIVA model expands on this theory and provides some evidence for the underlying neural processes (Guenther, 2016).

Some authors explicitly state that chunking is phonological/motoric, in which case phonemes are chunked together to make a consonant cluster, syllable, word or even phrase. These units are stored in memory and can then be retrieved/produced together as a group, instead of having to be formed on the fly. However, there is some evidence to suggest that chunking is partially semantic in nature. By this account, items which are semantically similar are stored together as a chunk in memory. For example, the CHREST (Chunk Hierarchy and REtrieval STructures) model developed by Gobet et al (2001) models learning as a network of nodes (chunks) which are connected in various ways. Learned information is stored in the form of these nodes in long-term memory (through a discrimination network). One of the ways a node might be connected to another node is by a lateral link, which refers to a semantic association between 2 nodes within the discrimination network. Therefore, associated semantic items are stored together in memory.

Finally, both children and adults show a processing advantage for the production of phrases which are high in frequency. Since many models of production follow a words and rules approach (e.g., Pinker and Ullman, 2002), they do not make specific predictions for this effect of phrase frequency. One interpretation is that frequent use of phrases also leads to chunking and that chunked multi-word sequences have a processing advantage (e.g., Bybee, 2006). In this account, frequent sequences are represented together as units that can be accessed directly rather than having to be formed.

The speed with which linguistic input is heard and with which words are produced imposes a severe constraint on the language system. In fact, speed of speech requires the language system to work faster than the memory system should allow for. Christiansen and Chater (2016) call this the Now-or-Never bottleneck. The authors propose that to deal with the Now-or-Never bottleneck, the language system undergoes the Chunk-and-Pass procedure.

During comprehension, information is coded immediately to ensure new that new incoming information does not overwrite it. The language system chunks this information over different levels of representations. For example, the signal might first be chunked at the phonological level, which then might be recoded into higher-level units like morphemes or words. These words might then be chunked into larger units like phrases. Finally, phrases might be recoded into higher-level discourse structures which may then be chunked further into an even more abstract representational structure. Once recoded, the information is no longer subject to interference from further auditory input. During this process, the system is supported by predictive knowledge taken from experience with sentences. This statistical information provides powerful predictive constraints on language comprehension.

During production, the process is reversed. First, discourse-level chunks are retrieved. These chunks are then broken down into their subsequent sub chunks of decreasing linguistic abstraction (e.g., phrases, words, phonemes). Once the system has retrieved chunks with sufficient information to drive the articulators, speech follows. This idea where there is a discourse-level representation of the intended message with lower-level chunks containing linguistic information is compatible with several current models of language production (e.g., Dell et al. 1997; Levelt 1998; 2001).

Although there is debate as to the nature or underlying processes of chunks, it is generally accepted that speakers do undergo some form of chunking to deal with the speed with which linguistic input is heard and the speed with which words are produced. In this thesis, we consider the processes which drive phonological/motor chunking during production in both children and adults, and whether this is a process which is used for phrases.

Finally, much of the evidence concerning lexical access and phonological/motor chunking comes from work with adults. While it is plausible that there is continuity between children and adults, results with adults cannot be unreflectively generalised to children. Children are still developing their conceptual knowledge, so it is possible that this knowledge is incomplete. We also might expect to see more between-word variability during the phonological level stage of production than is seen in adults, given that children are still practicing their production of words, while adults should have robust phonological knowledge of the words and phrases in their language. The aim of this thesis is to explore the effect of phrase frequency, word frequency and complexity (and their interactions) on production fluency in both children and adults, with a particular focus on any differences that exist between the two groups.

1.9 Structure of the thesis

The structure of the thesis will be as follows:

1.9.1 Chapter 2

Chapter 2 reports on the methodology used in the thesis. We first discuss the motivation for using sentence repetition tasks. We place particular focus on evidence for this being a robust method to explore the same underlying processes used during production as in spontaneous speech. We next discuss the coding of our experiments. We chose to code at the phoneme level using software which allowed us to also consider the speech wave. This meant that we were able to undertake more detailed error coding than coding based on audio files alone. We were able to see specifically where errors were being made which gave us more control over our analyses. We also coded the duration of productions using the same software. Considering the speech wave as well as the audio allowed for more precise coding. Finally, we explain the processes used in our statistical analyses.

1.9.2 Chapter 3

Chapter 3 reports on an experiment which explores the role of word frequency, phrase frequency and phonological complexity on fluency in children aged 3;2 to 3;9. To answer our research questions, the 3 independent variables were fully crossed. This is important because these 3 factors have not been fully crossed before. For example, experiments which consider phrase frequency effects tend to hold word frequency constant. Children were required to repeat phrases manipulated for these variables and both error rate and duration of production were measured. This allowed us to look at the effect of our independent variables (word frequency, phrase frequency, complexity), and their interactions, on measures of fluency (error rate and duration).

1.9.3 Chapter 4

Chapter 4 reports on the same experiment in children aged 3;9 to 4;4. Again, children were required to repeat phrases manipulated for these variables and we measured both error rate and duration of production. This allowed us to look for developmental changes in representation and processing.

1.9.4 Chapter 5

Chapter 5 reports on the same experiment in adults. Adults were required to repeat phrases manipulated for these variables and duration of production was measured. This allowed us to look for differences in representation and processing between children and adults.

1.9.5 Chapter 6

In chapter 6 we aim to rule out any effect of frequency being due to semantic related differences and to further explore the effects of our variables on novel stimuli. To do this, we explored the effect of word frequency and phrase frequency on the accuracy of phoneme production in adults using artificial stimuli. Again, word frequency and phrase frequency were fully crossed so that we could also explore their interaction. In experiment 4, adults were asked to listen to and then repeat non-words manipulated for frequency via a training session. In this experiment, we aimed to explore the effect of our variables on duration rate in adults for novel stimuli. In experiment 5, adults were presented with both the auditory and orthographic representation of the same non-word phrases. In this experiment, we aimed to explore the effect of frequency in an experiment more comparable to natural language - since the stimulus is displayed, the target phrase does not need to be held in working memory. The displayed stimulus acts as long-term memory to make this experiment more comparable to real language use.

1.9.6 Chapter 7 – General Discussion

Chapter 7 summarises the information found in the previous experiments and discusses in terms of the previous literature. The objective of this thesis was to explore the effect of phrase frequency, word frequency and complexity (and their interactions) on production fluency in both children and adults. We aimed to explore whether children's phonological learning, like that of adults, involves phonological/motor chunking. The results from these experiments demonstrate that young children and adults do engage in phonological/motor chunking, and that this effect is driven primarily by word frequency. We did not find any effect of phrase frequency in this thesis which might suggest that previously reported effects of phrase frequency are not due to phonological/motor chunking, but to some other factor (e.g., redintegration via long term memory).

2 Chapter 2 – General Methods

The aim of this chapter is to describe the methods employed in this thesis, and why they were chosen. We will start by exploring the processes underlying production in sentence repetition tasks, which is the task used to explore the research questions in this thesis. We will then describe the coding process and statistical analysis used in the experimental chapters.

2.1 Sentence repetition tasks

Sentence repetition tasks are typically used to measure language ability in children. Poor performance on this task has been argued to be a useful clinical marker of Developmental Language Disorder (DLD, e.g., Conti-Ramsden, Botting, & Faragher, 2001). However, the paradigm has also been widely used to explore the representations and processes underlying production in typically developing children and even in adult speakers. Our use of the method is in line with this second body of work.

In the following chapters, we use a series of elicited imitation tasks to explore the effect of phrase frequency, word frequency and phonological complexity on fluency. In our experiments, children and adults were required to repeat phrases manipulated for these variables. Their subsequent productions were then measured for error rate and duration of production, which we take as measures of fluency. We use this sentence repetition task to explore the processes which underlie production. Before we describe these experiments, however, we will review the previous literature examining which underlying processes this task measures.

2.1.1 Sentence repetition and working memory

Working memory is the part of the short-term memory system which deals primarily with processing and temporarily storing information. The most widely-cited model of working memory was developed by Baddeley and Hitch (1974) who proposed that there are 3 main components to working memory. The first of these is the central executive which controls the stream of information between the phonological loop (which stores verbal/phonological information) and the visual-spatial sketchpad (which stores visuo-spatial information). In this model, long-term memory was thought to be a distinct system. However, evidence for long-term memory influencing recall has since emerged and Baddeley (2000) revised this model to include the episodic buffer. This is a limited capacity system which links the above components and long-term memory.

This model can be useful in understanding what is happening in a sentence repetition task. We will sketch how this might work. In a sentence repetition task, the speaker is first required to retain and recall the target sentence. It is possible that the episodic buffer supports this short-term retention during sentence repetition by integrating representations of words/phrases (from working memory and long-term memory). Where there are strong representations of words and phrases, retention will be well supported. For novel phrases, there will be no representations available. In this case, the episodic buffer relies on knowledge of phonemes, and retention is more difficult. By this account, chunking/frequency effects are rooted in the episodic buffer. Next, the speaker uses the articulatory system to produce the target sentence (there is debate as to how this production system might work – see section 1.6).

The regeneration hypothesis aims to describe the mechanisms that speakers use during sentence repetition tasks (Potter & Lombardi, 1990). This hypothesis contends that long-term memory plays a primary role and is based on a production model proposed by Bock and Levelt (1994). The authors propose that to repeat a sentence, the speaker is first required to retrieve the conceptual representation of the target word or phrase from long term memory, following which sentence production proceeds as in free speech production (e.g., moving words into the syntactic structure, completing morphological units, etc.). This model involves all aspects of the memory system, as well as the language production system.

Initial evidence for this hypothesis was taken from a series of experiments conducted by Potter and Lombardi (1990) who asked speakers to read or listen to a sentence and then (after a brief distractor task) recall the sentence. It was found that when a synonym for one of the words in the sentence was included in the distractor task (e.g., castle for palace), intrusion of that word during recall was more frequent. The authors took this finding as evidence for the speaker expressing the conceptual representation in a way much like in spontaneous speech production, through processes in long-term memory where lemma-level representations are found. The authors use the term 'reconstruction' to describe the processes underlying these semantic errors, where the lexical selections are chosen from the most recently active units reassembled from information in long-term memory (syntactic and semantic information), even when the short-term memory trace has completely decayed (see also Lombardi and Potter, 1992). Therefore, sentence repetition tasks involve the construction of the sentence from long-term memory representations.

Another theory that has been used to account for sentence repetition is called redintegration. This theory is like the reconstruction hypothesis but instead views long-term memory as supporting representations in short-term memory. Here, long-term memory processes enable the organisation of information into larger units or 'chunks.' Jacobs, Dell and Bannard (2017) suggest that redintegration drives phrase frequency effects during recall. According to their account speakers store phrasal representations consisting of the constituent lemmas, which are held together by long-term memory associations. The lemmas cue each other based on how often they have co-occurred, which means that the memory system can fill in the gaps for a phrase when not all words are retrieved. It is this process that leads to frequent sequences being retrieved more efficiently than infrequent sequences. The authors found that speakers were more likely to recall both words from high frequency phrases given recall of at least one word. This suggests that once one word is retrieved from the phrase, the phrase is more likely to be completed in its entirety for high frequency items. Therefore, redintegration is driven by long-term memory associations between lemmas of to-be-recalled items, where the entire phrase is more likely to be recalled given recall of one lemma.

Riches (2012) also provides evidence suggesting that redintegration plays a role in sentence repetition. This experiment aimed to explore which mechanisms are used during sentence repetition tasks in children who are typically developing and in children with DLD. The reconstruction hypothesis contends that sentence repetition involves the total reassembly of the sentence from representations stored in long-term memory. Riches (2012) argues that if the processes of reconstruction do stem from long-term memory alone, there will be no observed latency effect during sentence repetition - because information in long term memory should undergo a slow rate of degradation. However, recall in the DLD group showed a greater effect of latency than the age matched group, which suggests the involvement of short-term memory (which has a rapid rate of degradation). The author argues that this result is in line with redintegration, where long-term memory supports representations held in short-term memory.

Given the above findings, we argue that participants performing a sentence repetition task make use of at least some of the representations and processes utilised in spontaneous speech. Therefore, we chose to use sentence repetition tasks to explore our research questions which are concerned with the nature of those representations and processes.

2.2 Coding and data analysis

For each experiment, the productions of the participants were recorded. These recordings were then coded in full by the author and a subset of the data was second coded by another researcher. To answer our research questions, we measured both the error rate of phoneme production and duration of production, before analysing the data using multilevel modelling. The following section aims to describe the techniques involved in the coding and analysis of data.

2.2.1 Coding for error rate (phoneme by phoneme)

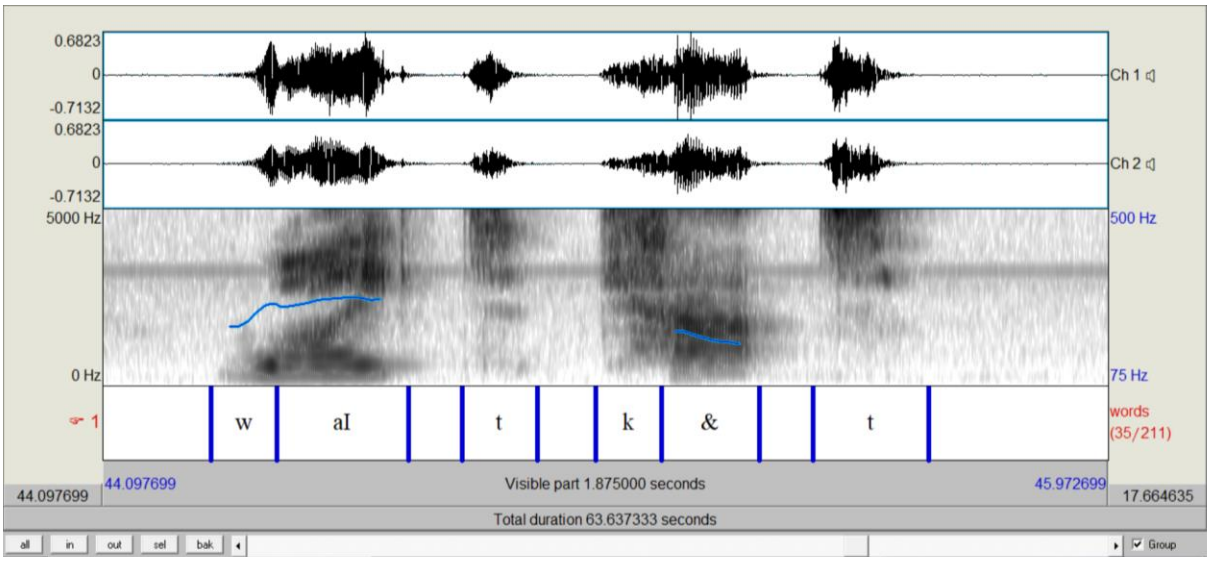
The coding of error rate in sentence repetition data in the past has largely focused on the coding of audio files (e.g., Dell, 1990) where, in most cases, the entire word or phrase is coded to be correct or incorrect and an accuracy score (the proportion of correct productions overall) is based on this. For example, Sasisekaran et al. (2010) asked children aged 9–10 years and adults to produce non-words over 2 sessions in an elicited imitation experiment. The authors tracked articulatory movement and audio recorded the productions so that the experimenter also had access to the face and articulators of the speaker. Correct responses were defined as those trials which included no articulation or disfluency errors. From this, the number of correct responses was divided by the number of repetitions of that word to calculate the percentage of correct productions for the entire word, which was used as the measure of accuracy.

Bannard and Matthews (2008) tested 2- and 3-year-old children’s knowledge of frequent multiword sequences via a repetition task using pairs of sequences that were identical except for the final word. In this experiment, both error rate and duration were used as measures of fluency. For accuracy, the authors used a similar methodology to that explained above. Again, each word was coded to be correct or incorrect. If a child did not make a single error in an entire sequence, this sequence was coded as correctly repeated. The authors next calculated the mean proportion of correctly repeated sequences and compared for high frequency phrases and low frequency phrases.

To measure the error rate of speaker’s productions, we chose to code productions phoneme by phoneme. We chose to use phoneme-by-phoneme coding because this gave us a more detailed coding. This allows one to identify specifically where errors are being made, opening the possibility of understanding which types of individual phonemes are most likely to be correctly produced. Furthermore, if you are interested in the properties of words and phrases rather than phonemes directly it can capture degrees of accuracy - the proportion of component phonemes that were accurately produced.

To perform this coding, recordings were opened in Praat (Boersma & Weenink, 2018) where the children's phoneme productions were transcribed using the CELEX transcription alphabet, which is a machine-readable adaptation of the International Phonetic Alphabet. To determine each phoneme, both the waveform and spectrogram were used to guide coding. For instance, you can see in Figure 2-1 that the phonemes 't' and 'k' are characterised by a burst in intensity in the spectrogram. The start and end of the production are also clear to see from the waveform and spectrogram. For each phoneme, shorter segments of the waveform were examined and listened to in isolation. As you can see in the example transcription below, the coding indicates that the speaker has produced the target phrase 'white cat' correctly.

Figure 2-1 Example transcription of 'White Cat'



This coding was extracted and each of the target phonemes were given a score of 0 or 1 depending on whether a match was found for that phoneme in the child's attempted production when the two sequences (target phoneme sequence and child's produced phoneme sequence) were aligned using the Needleman-Wunsch Algorithm (Needleman & Wunsch, 1970).

This algorithm is often used to compare biological sequences (e.g., amino acids). The algorithm assigns a score to every possible alignment of two sequence - in our case the target sequence and the sequence produced. Points are given for each match and penalties are applied for each substitution, insertion, and deletion in the alignment. As is standard we set the penalty for insertions and deletions to be the same as that of a substitution.

The alignment with the highest score is chosen and output. The advantage of this fully automated process is that it is transparent - making the assumptions about which alignment out of all the possible alignments should be preferred, as stated above - and replicable. Below, are examples of the alignments chosen by the algorithm containing deletions, substitutions, and insertions.

Table 2-1 Example of deletion:

Phrase	Word	Operation	Target Phoneme	Phoneme Produced
Loud Bear	Loud	MATCH	l	l
Loud Bear	Loud	MATCH	aU	aU
Loud Bear	Loud	DELETION	d	^
Loud Bear	Bear	MATCH	b	b
Loud Bear	Bear	MATCH	E@	E@

Target phonemes for the phrase ‘loud bear’ are on the left while phonemes produced are on the right. As you can see, the speaker did produce the target phoneme ‘l.’ The speaker also produced the target phoneme ‘aU.’ These are matches. However, the speaker did not produce the target phoneme ‘d.’ This is a deletion.

Table 2-2 Example of substitution:

Phrase	Word	Operation	Target Phoneme	Phoneme Produced
Bad Car	Bad	SUBSTITUTION	b	p
Bad Car	Bad	MATCH	&	&
Bad Car	Bad	MATCH	d	d
Bad Car	Car	MATCH	k	k
Bad Car	Car	MATCH	A:	A:

In the above example, the speaker did not produce the target phoneme ‘b’ instead the speaker produced the target phoneme ‘p.’ This is a substitution.

Table 2-3 Example of insertion:

Phrase	Word	Operation	Target Phoneme	Phoneme Produced
Pink Star	Pink Star	INSERTION	^	s
Pink Star	Pink Star	MATCH	p	p
Pink Star	Pink Star	MATCH	I	I
Pink Star	Pink Star	MATCH	N	N
Pink Star	Pink Star	MATCH	k	k
Pink Star	Pink Star	MATCH	s	s
Pink Star	Pink Star	MATCH	t	t
Pink Star	Pink Star	MATCH	A:	A:

In the above example, the speaker did not produce the target phoneme ‘p’ instead the speaker produced the target phoneme ‘s.’ This is an insertion.

To conduct the statistical analyses for these error data, we first removed insertions from the dataset. All matches were then given a score of 1 and all other operations were given a score of 0. The accuracy score (0 and 1s) was used directly in a logistic regression. Finally, we built a series of multilevel regression models and used likelihood ratio tests in an iterative drop-one analysis to find the model with best fit (see section 2.3).

2.2.2 Second coding for error data

2.2.2.1 Child data (Chapter 3 and Chapter 4)

20% of the recordings were separately coded by a second researcher and the same alignment process performed. The second coder’s coding was then compared to the coding of the author. Agreement statistics are reported for each study. We found that there were some common disagreements concerning particular phonemes. In some cases, it is possible that these disagreements are driven by regional differences in accents. We also found that phonemes which sound very similar were sometimes confused. For example, it was common for coders to disagree as to whether the child was producing a ‘t’ or a ‘d’. As you can see from Table 2-4, these phonemes are, according to a canonical phonological description, identical on manner features and place features. It was also common for coders to disagree as to whether the child was producing a ‘k’ or a ‘g.’ Again, these phonemes are identical on manner features and place features.

Table 2-4 Consonants single place of articulation Table (p. 108) in Introductory Phonology (Hayes, 2009):

	Manner features								Laryngeal features			Place features													
	consonantal	sonorant	continuant	delayed release	approximant	tap	trill	nasal	voice	spread gl	constr gl	labial	round	labiodental	coronal	anterior	distributed	strident	lateral	dorsal	high	low	front	back	tense
t	+	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	-	-	-	-	0	0	0	0	0
d	+	-	-	-	-	-	-	-	+	-	-	-	-	-	+	+	-	-	-	-	0	0	0	0	0
k	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	0	0	-	+	+	-	0	0	0
g	+	-	-	-	-	-	-	-	+	-	-	-	-	-	-	0	0	0	-	+	+	-	0	0	0

To explore the consequences of these disagreements for our analyses we first identify all specific phonemes (in specific word contexts) where there were disagreements that affected the correctness score given – phonemes which had been coded to be correct by one coder (the phoneme matched that of the target phoneme), and as a different phoneme by the other coder. For example, in Table 2-5 you can see that the phrase ‘pink star’ was coded to be correct by coder 1, while coder 2 disagreed on the coding of ‘t’ and instead coded ‘d’.

Table 2-5: Example of disagreement between ‘k’ and ‘g’

Phrase	Word	Target Phoneme	Coder 1	Coder 2
Pink Star	Pink Star	p	p	p
Pink Star	Pink Star	I	I	I
Pink Star	Pink Star	N	N	N
Pink Star	Pink Star	k	k	k
Pink Star	Pink Star	s	s	s
Pink Star	Pink Star	t	t	d
Pink Star	Pink Star	A:	A:	A:

This allows us to conduct a secondary analysis to look at whether our results are contingent on a coding decision about which there is disagreement. When a target phoneme was coded as a different phoneme for at least two of the children, we allowed this different phoneme to be considered as an alternative form of the target for the purposes of this secondary analysis. This meant that we allowed a match for both the main coder and the second coder, where that coder had said that the child produced the permitted alternative form.

We then re-ran the relevant analyses with the alternative dataset to see how this new matching impacted results. Where the alternative dataset gives us the same results, we take this as evidence that the effects are robust to the coder differences. More details are provided for each experiment where this is performed.

2.2.2.2 Non-word coding (Chapter 6)

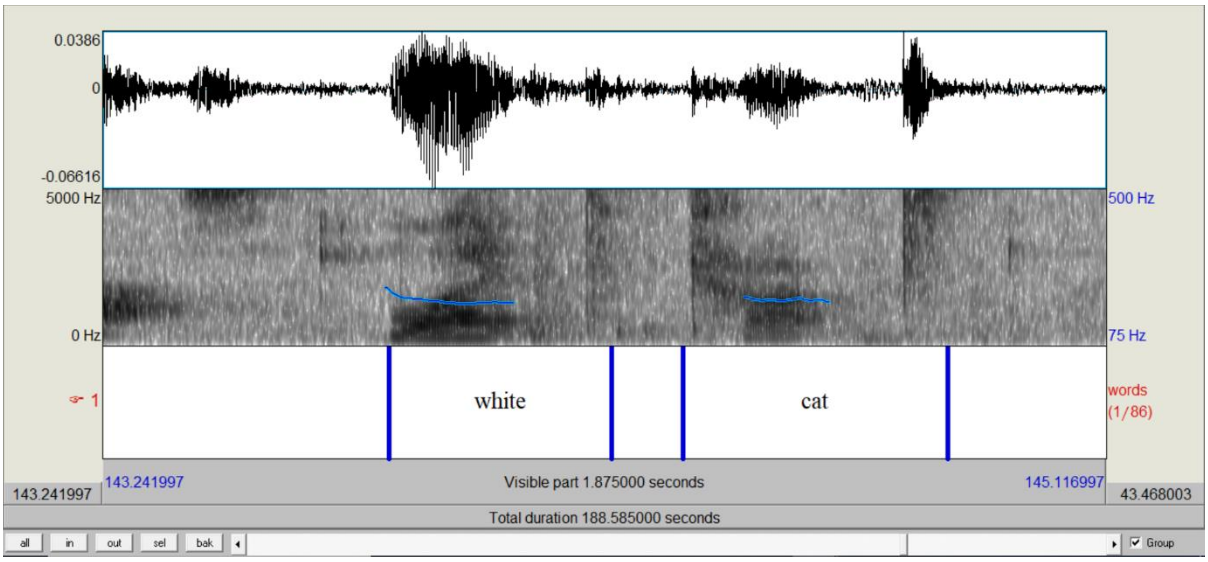
The speech waves for the non-word data were also additionally coded by a second researcher. Again, the second coder was asked to identify phonemes produced in the same way as the author. The author’s coding was compared with this second coding and Kappa scores were calculated and reported. No secondary analysis was performed here because Kappa scores indicated ‘substantial’ agreement (Cohen, 1960).

2.2.3 Coding for duration data

For the child data, the length of the coded phonemes (see above example of white cat – section 2.2.1) were extracted from the phoneme-by phoneme coding in Praat. The individual phoneme durations were taken and combined to produce a word duration. The word frequency and complexity scores for the component words of each target phrase are nested within the target phrase (each word has a different word frequency and word complexity score although they are in the same frequency/complexity band). Therefore, we chose to calculate the length of production duration for words. The duration of each word was then combined to give phrase duration. To do this we chose to commute phone lengths rather than considering the onset and offset of durations. This is because we were interested in the time taken to perform the specific target motor routines and not the overall time taken including silence. However, it is possible that this was a consequential decision – in future research we could consider analysing the data each way (summing of phone lengths and considering onset and offset durations).

For the adult repetition data, the duration of each word was coded. To do this, recordings were opened in Praat and the onset and end of each word was identified. The reason this different approach was taken for the adults is that since the phoneme-by-phoneme coding was not available, we took the finest details information available.

Figure 2-2 Example transcription of 'White Cat'



To conduct the statistical analysis, we first removed phrases that were not produced correctly from the dataset. Next, we built a series of multilevel linear regression models and used likelihood ratio tests in an iterative drop-one analysis to find the model with best fit.

2.3 Statistical analysis

2.3.1 Hierarchical data structures in repeated measure designs

The experiments in this thesis use a repeated measures design, where multiple observations are made within the same individual. Each participant was required to repeat multiple phrases manipulated for our independent variables. This results in nested hierarchical structures where datapoints are non-independent. To account for this nesting, multilevel modelling was conducted.

2.3.2 Multilevel regression models

This analysis models clustering by partitioning the overall variance into separate levels, which allows for the predictors of both within and between differences (and their interactions) to be modelled (Szmaragdand & Leckie, 2013). The simplest multilevel model is a 2-level random intercept model. This could be used in an analysis where you expect observations to be nested only within participants. In this case, the inclusion of participant as a random intercept allows for between-person differences in the mean proportion of the dependent variable. In our analyses, we included participant-varying intercepts. However, we also included random effects of phrase, and position (part of speech - adjective or noun - in the real language experiments and first or second position in the artificial language experiments) as random intercepts. This means that the intercepts vary among these units. In addition to this, we included a random effect of participant on all slopes. This models participant variance in the effect of each of our predictors on responses (e.g., it allows for predictors to have a positive association with production for some participants, and a negative association for others) while allowing estimation of a population-level slope. It thereby allows us to infer that any population-level effect seen would generalise to a new sample.

2.3.3 Model selection

To determine the value of each of our independent variables and decide on a model to report for each of our studies we perform the following procedure. We started by building a full model including all our independent variables and their relevant 2-way interactions as fixed effects. Random effects of participant, part of phrase (POS - adjective or noun) and item (phrase) on the intercept and a random effect of participant on all slopes were also included in the models. We then used likelihood ratio tests in an iterative drop-one analysis. We first looked at the effect of any pairwise interactions by building a full model and removing each interaction separately. Having removed any non-significant interactions but keeping all individual variables, we then perform a drop-one analysis on the remaining terms. We then remove any non-significant variables and repeat until only variables that explain significant variance, along with any component terms where interactions are included remain. Finally, to look at the contribution of the individual variables, we remove all remaining interactions where appropriate and perform a drop-one analysis on a model containing only the individual model terms. When fixed effect terms were removed at each step, the random effects were left in for that term.

For example, when comparing the full 2-way interaction model to the reduced comparison model the following models were used:

Full 2-way interaction model:

CORRECT ~ (1 + WORDFREQ*PHRASEFREQ + WORDFREQ*WCM + PHRASEFREQ*WCM |PARTICIPANT) + (1|POS) + (1|PHRASE) + WORDFREQ*PHRASEFREQ + WORDFREQ*WCM + PHRASEFREQ*WCM)

Reduced model evaluating WF*PF:

CORRECT ~ (1 + WORDFREQ*PHRASEFREQ + WORDFREQ*WCM + PHRASEFREQ*WCM |PARTICIPANT) + (1|POS) + (1|PHRASE) + WORDFREQ*WCM + PHRASEFREQ*WCM)

Here the WF*PF interaction is removed from the fixed effects, but random effects for the full 2-way interaction are still included. If a likelihood-ratio test showed that doing this did not harm fit, the fixed term is assumed to explain significant variance.

3 Chapter 3 - Assessing the effect of word frequency, phrase frequency and phonological complexity on fluency in 3-year-olds

3.1 Introduction

Word frequency has been shown to affect production in both children and adults, where speakers are more fluent (are faster and/or make fewer error) in producing words that occur more frequently in the language. Sosa and Stoel-Gammon (2012) explored the production of monosyllabic target words in 15 children aged 2;0 and 2;5 and found that although words that were more phonetically complex (containing later developing consonants and syllable structures) were produced more variably, high frequency words were less variable than low frequency words. Similarly, Dell (1990) showed that during a word naming task, adults made fewer phonological errors when producing more frequent words for both function and content items.

Different models of word production have sought to explain this effect. Although there are many diverse models of speech production (e.g., Garrett, 1975; Levelt, 1989; Caramazza, 1997), most models contend that the production of words requires 2 stages. For example, Levelt's (1989; 2001) two-stage lexical access model proposes that during the first stage (called conceptualisation) the speaker first specifies a concept, which is then represented as a linguistic form (lemma). Next, during a second stage (called formulation), activation spreads from the lemma to the relevant phonological items, which make up the phonological form of the word. Finally, articulation takes place during this formulation stage, which involves the motor execution and production of the word.

Word frequency effects in adults are most commonly attributed to the second (formulation) stage of lexical access, where the phonological form of the word is realised (see Levelt, Roelofs & Meyer, 1999). In support of this, Jescheniak & Levelt (1994) found that speakers produced homophones at the same speed as control words matched for the combined frequency of both homophones, which was faster than their production of control words matched for independent word frequency. The authors took this to mean that the observed frequency effects are contained at the phonological level, given that homophones have a shared level of representation during the formulation stage. In addition to this, both adults (Harley & MacAndrew, 2001) and aphasic patients (Nickels, 1995) have been shown to make more phonological, but not semantic, errors for words which are low in frequency compared to words which are high in frequency. It is generally accepted that semantic errors are associated with the first conceptualisation stage of lexical access, while phonological errors are associated with the second formulation stage of lexical access. Given that the rate of semantic errors is not affected by frequency, it was argued that frequency does not have an effect during the conceptualisation stage.

There is, however, some other work that does suggest that word frequency affects lexical selection at the semantic level. Kittredge, Dell, Verkuilen and Schwartz (2008) did find that aphasic patients made fewer phonological (e.g., producing ‘pillow’ instead of ‘pineapple’) and semantic errors (e.g., producing ‘apricot’ instead of ‘pineapple’) in high frequency word targets during a picture naming task, although this frequency effect was stronger for phonological errors. The authors argue that these results demonstrate that frequency has an effect at both the conceptualisation and formulation stage, and not only at the formulation stage. Navarette, Basagni, Alorio and Costa (2006) found that adult’s responses were faster for pictures with high-frequency names compared to low-frequency names in a gender decision making task that only required lemma retrieval. In experiment 2, speakers were asked to determine whether the name of the target picture was masculine or feminine using a button press. The authors argue that this task requires lemma retrieval but does not require the processes involved in the formulation stage because the speaker does not produce the target name.

The evidence concerning the level of representation involved in errors comes from work with adults. While it is plausible that there is continuity between children and adults, results with adults cannot be unreflectively generalised to children. Healthy adults are assumed to have well-established conceptual knowledge, and robust knowledge of core vocabulary, such that the lemma-retrieval part of word production is assumed to be unproblematic. However, children are still developing their conceptual knowledge, and we might expect their knowledge to be incomplete. Hence, we might expect to see more between-word variability at that stage of production than in adults. Work is required to clarify the cognitive representations and mechanisms that produce word frequency effects in children and to determine whether these differ from adults.

Alongside the word frequency effects discussed above, the last decade has seen a number of papers published showing that speakers are affected not just by the frequency of words, but also of phrases, such that children and adults are better at producing high frequency phrases. Bannard and Matthews (2008) matched frequently occurring chunks in language (e.g., a piece of cheese) to infrequent sequences (e.g., a piece of food), where the final words (e.g., cheese, food) and final bigrams (of cheese, of food) were matched for frequency. It was found that 2- and 3-year-old children repeat high-frequency word sequence combinations more rapidly and accurately than low-frequency combinations. Similarly, Janssen and Barber (2012) found that native Spanish speaking adults took less time to produce more frequent adjective-noun and noun-noun phrases compared to the less frequent counterparts (object name frequency was held constant) in a picture naming task. It was also shown that native French speakers produced more frequent adjective-noun and determiner-adjective-noun phrases more quickly than less frequent phrases.

So far, models of production have focused on words and tend to follow the words and rules approach for phrases (e.g., Pinker and Ullman, 2002) where sentences are constructed by combining memorised word forms taken from the lexicon via knowledge of innate rules. However, the number of reported phrase frequency effects have led to the idea that frequent use of phrases leads to chunking, where chunked multi-word sequences have a processing advantage. One interpretation of this is the usage-based model (Bybee, 2006), which contends that frequent sequences are represented together as units which can be accessed directly rather than having to be formed. Despite this, it is less clear at which stage of lexical access phrase frequency effects might occur. Furthermore, there is essentially no understanding of how this might change across development. Work is required to clarify the cognitive representations and mechanisms underlying phrase frequency that affect production in children, and to further determine whether these differ from adults.

In this research we are interested in how frequency influences production in children. Assuming that errors do come from the formulation stage, frequency-related effects in adult production are usually conceptualised, even if only implicitly, as an effect of phonological/motor sequence learning, with words that are difficult to articulate when first encountered becoming easier to say due to practice. Segawa et al. (2015) explored the trade-off between phonological complexity and frequency by asking American English speakers to produce mono-syllabic non-words containing phonotactically legal (*e.g., BLERK, THRIMF, TRALP*) and illegal (*e.g., FPESHCH, GVAZF, TPIFF*) bi- or triconsonantal initial and final consonant clusters. It was found that speakers had a significantly lower error rate after practice for the illegal sequences but did not differ in their repetition of legal syllables. Furthermore, after introducing novel illegal sequences, it was found that speakers made more errors during novel illegal syllables compared to learned illegal syllables. This latter finding provides evidence for learned motor programs - given that performance improvement was specific to the stimuli encountered, improvement could not be a result of the learning of phonological rules. Therefore, speakers improve in their production of complex non-words by engaging in some kind of phonological/motor chunking.

The aim of the current experiment was to explore the effect of phrase frequency, word frequency and word complexity (and their interactions) on the accuracy of phoneme production for phrases during an elicited imitation task. We are interested in whether, like for adults, frequency effects in children are attributable to phonological/motor chunking. Here, the production system is said to store frequent phoneme sequences as a single ‘item’ – or chunk. This means that the phonological form of the item is activated more quickly than for targets which are not stored as a chunk. In addition to this, the motor routines for these sequences are also more efficient due to repeated production practice (e.g., Segawa et al., 2019). However, it is also possible that frequency effects might plausibly be attributed to another stage of production. To explore this, participants were asked to repeat 24 adjective-noun phrases which were manipulated for phrase frequency, word frequency and word complexity. Their productions were coded phoneme by phoneme.

If motor learning does occur in 3-year-olds, words should become easier to say after practice. We might expect children to make few errors when repeating words and phrases which are low in complexity, regardless of frequency. This is because low complexity items do not need practice to overcome their inherent articulatory complexity. For complex words and phrases, however, we expect children to make fewer errors when producing phrases that are high in frequency or contain high frequency words, compared to complex phrases which are low in frequency or contain low frequency words. These outcomes would provide evidence of phonological/motor chunking, whereby motor routines have become more efficient after practice.

It is also important to note that this is the first study to cross factors of word and phrase frequency. We further aim to explore whether these effects are independent and, if not, whether they are complementary or if the accessing of representations at different levels might lead to competition. Sosa and Macfarlane (2002) found that adults took longer to respond to the word ‘of’ for collocations (e.g., ‘part of’) when the collocation was higher in frequency. The authors took this result as evidence of competition between the representation of the word and of the phrase so that reaction times were faster for phrases which were more predictable but were slower when there was competition between the particle and ‘chunked’ collocation.

We therefore test the following hypotheses:

1) Children will find it easier to repeat phonemes in low complexity words.
2) Children's ability to repeat phonemes in phrases will be affected by the frequency of the phrases and their component words.
3) The negative effect of WCM (word complexity) on children's ability to repeat phonemes contained in each word will be reduced when that word is high frequency or part of a high frequency phrase.

We also test the secondary prediction that phrase and word frequency will affect (promote or suppress) each other's effect on accuracy.

3.2 Method

This study was preregistered on the Open Science Framework website (https://osf.io/9k4jn/).

3.2.1 Participants

65 monolingual English speaking 3-year-olds who were aged between 3;2 and 3;9 (mean age 3;6) were recruited and included in analyses. Consent was given and the study was approved by the ethics board. All children were tested in preschools and nurseries in Bedfordshire and Buckinghamshire. 15 additional children were tested but were excluded. 6 children refused to take part in the experiment and 7 children were replaced because they failed to produce at least 18 codable responses (75% of the items presented) as planned in the study preregistration. 1 further child with autism was tested but their data not included in analyses.

A sample size of 65 was determined in advance (and preregistered) based upon a bootstrap power analysis using a pilot sample of 30 participants collected using a similar and identically sized stimuli set. The model used was a multilevel logistic regression fitted in R using the package lme4. All independent variables were included but no interactions, meaning that the study is powered for the effect of these variables alone. Participant- and part-of-speech-varying random intercepts were included but no random slopes (as they did not improve fit on the pilot sample). 65 was the minimum number of participants at which 0 was not contained in the 95% bootstrapped intervals (based on 10000 random samples at each sample size) for any of the three independent variables (word complexity, word frequency and phrase frequency).

3.2.2 Design

This study used a within-subjects design. Our dependent variable was a score of 1 (correctly produced) or 0 (not correctly produced) for each phoneme in the target items. See section 2.2.1 for more details.

Our independent variables were as follows:

- Log Phrase Frequency
- Log Word Frequency
- Word Complexity

3.2.3 Materials

27 English adjective-noun combinations (combinations of monosyllabic adjectives and monosyllabic nouns) found in a corpus consisting of all mothers' or fathers' speech in English language transcripts in the Child Language Data Exchange System (CHILDES) database were used (MacWhinney, 2000). Frequencies were taken from the combined set of all transcripts. Three of these phrases were practise items. The non-practise phrases were chosen so that they varied in phrase and word frequency in the transcripts, as well as word complexity. Although all predictors are treated as continuous variables in our analysis, for the purposes of item selection, banding was used. The band cut-off points were chosen for each predictor, given the observed distribution, with the goal of achieving maximum spread,

but also allowing the different predictors to be crossed and decorrelated as much as possible.

3.2.3.1 Phrase frequency:

Phrases with a Log Frequency in the top 6/7ths of the range of adjective-noun phrases extracted from the corpus were considered to be high frequency. So, phrases with a Log Frequency greater than 1.195229. Range: 3.871201 (‘white cat’) - 1.94591 (e.g., ‘fun toy’). Phrases in the bottom 1/7th were considered to be low frequency. So, phrases with a Log Frequency less than 1.195229. Range: 1.098612 (‘pink star’) – 0 (e.g., ‘cold cream’).

3.2.3.2 Word frequency

Words with a Log Frequency in the top 2/5ths were considered to be high frequency. So, words with a Log Frequency score greater than 6.541476. Range: 9.803999 (‘big’) - 6.595781 (‘house’). Words in the bottom 3/5ths were considered to be low frequency. So, words with a Log Frequency score less than 6.541476. Range: 6.51323 (‘noise’) - 3.433987 (e.g., ‘hill’).

3.2.3.3 Word complexity measure (WCM)

Word complexity was operationalised using the Word Complexity Measure developed by Stoel-Gammon (2010). Emphasis is placed on the complexity of sound classes, syllable structures and word patterns produced (see Table 3-1), where higher scores reflect more complex/later developed phonology. The WCM outputs a number based on phonetic complexity of productions. A score is awarded to each word based on the presence of each of the features shown in Table 3-1. In this case, words with a complexity score of 3 or more were considered to be of high complexity, whereas those with a score of 2 or less were considered to be low complexity, given the higher rate of words with a measure of 2 or less (number of nouns: score 0-2 = 18,582, score 3-11 = 19,763; number of adjectives: score 0-2 = 17,526, score 3-13 = 20,819).

Table 3-1 Word Complexity Measure

Factor	One point awarded
Word patterns	Stress on any syllable but the first
Syllable structures	Word-final consonant
	Consonant cluster (sequence of two or more consonants within a syllable)
Sound classes	Velar consonant
	Liquid, syllabic liquid, rhotic vowel
	Fricative, Affricative
	Voiced fricative or affricative

*One point given for each sound class in production

There was no significant correlation (across the set of 24 items without repetitions; all correlation coefficients are below .3) between any of these independent variables, thus maximising the chance of attributing unique variance to the predictors. Three phrases were used for each of the 8 unique combinations of levels for the binary factors. See Table 3-2 for details of the 24 target items. 3 practise phrases were also included.

Table 3-2 Target items with Frequency/Complexity Bands

Phrase	Phrase Frequency		Word Frequency			WCM score		
	Band	Log	Band	Adjective Log	Noun Log	Band	Adjective	Noun
Old house	High	3.688879	High	7.915348	6.595781	High	3	4
Green square	High	2.944439	High	8.791182	6.872128	High	4	6
Sweet girl	High	2.9957323	High	7.01301579	8.79769958	High	4	4
White cat	High	3.871201	High	8.040447	6.747587	Low	2	2
Big room	High	2.70805	High	9.803999	6.761573	Low	2	2
Fun toy	High	1.94591	High	7.783224	7.640604	Low	2	0
Steep hill	High	1.94591	Low	3.433987	3.828641396	High	3	3
Strange noise	High	2.197225	Low	5.926926	6.51323	High	6	3
Slow train	High	1.386294	Low	5.916202	6.324359	High	3	3
Deep sea	High	1.386294	Low	5.613128	3.526361	Low	1	1
Short tail	High	1.94591	Low	6.142037	5.541264	Low	2	2
Sore mouth	High	2.197225	Low	6.308098	4.94876	Low	2	2
Pink star	Low	1.098612	High	7.808729	6.891626	High	4	3

Cold cream	Low	0	High	8.105911	7.400621	High	4	4
Cute box	Low	0	High	6.980076	6.807935	High	3	4
Loud bear	Low	0	High	6.665684	7.198184	Low	2	1
Bad car	Low	0	High	7.628031	7.216709	Low	1	2
Wet head	Low	0	High	7.576097	8.597297	Low	2	2
Grey skirt	Low	0	Low	5.894403	3.433987	High	3	5
Gold watch	Low	0.693147	Low	5.713733	5.420535	High	4	3
Burnt milk	Low	0	Low	4.442651	5.497168	High	3	4
Thin line	Low	0	Low	5.17615	5.308268	Low	2	2
Odd shape	Low	0.693147	Low	4.682131	5.087596	Low	1	2
Cheap shop	Low	0	Low	3.912023	5.347108	Low	2	2

Table 3-3: Information about the raw frequencies of target words and phrases within frequency bands

HIGH PF 48 to 7	HIGH WF 18106 to 732	LOW PF 2 to 1	HIGH WF 5417 to 785
HIGH PF 9 to 4	LOW WF 5417 to 785	LOW PF 2 to 1	LOW WF 363 to 31

*The highest phrase frequency for possible adjective noun phrases was 4301 and the lowest was 1 while the highest word frequency for possible adjective or nouns was 54310 and the lowest was 1

Phrases were recorded by a speaker from Bedfordshire using stress on the first word. All phrases were counterbalanced.

3.2.4 Procedure:

This experiment used an elicited imitation paradigm (like that used in Bannard and Matthews, 2008). The experimenter sat with the child in front of a laptop and explained that they were going to play a game. The child was given a picture and told that they would cover the picture with stickers. It was explained that to get the stickers, they would need to listen to what the computer says and then repeat the same phrase. The experimenter then offered to go first and played the first example phrase. She repeated the sequence and then awarded herself a sticker. The experimenter then played the remaining 2 practise phrases. If the child did not respond within approximately 6 seconds, the experimenter prompted the child, and reminded them that they need to repeat what they hear. Each practice phrase was replayed if necessary. Every time the child attempted to repeat the phrase, he/she received a sticker. The test sequences were played next. For these experimental trials no help was given to the child and no phrases were repeated. If the child did not immediately respond, the experimenter waited for 8-10 seconds before asking “Can you say that?” The procedure continued until all 24 test phrases were produced.

3.2.5 Transcription and coding

The recordings were coded by the author. Recordings were opened in Praat (Boersma & Weenink, 2018) where the children’s phoneme productions were transcribed using the CELEX transcription alphabet, which is a machine-readable adaptation of the International Phonetic Alphabet. Next each of the produced sequences was aligned with their respective target sequences using the Needleman-Wunsch Algorithm. Each child was given a score of 0 or 1 for each phoneme in the target depending on whether a match was returned for that phoneme in the alignment (see section 2.2.1). 20% of the recordings were second coded by another researcher (see section 3.3.1 below). To check that the results that we report are not contingent on coder decisions about which there was not agreement, we also performed a robustness check as described in section 2.2.2.1.

3.3 Results

To address our research questions, we built a series of multilevel logistic regression models. We started by building a model including phrase frequency, word frequency, word complexity and their relevant 2-way interactions as fixed effects. Random effects of participant, part of phrase (adjective or noun) and item (phrase) on the intercept and a random effect of participant on all slopes were also included in the models. Our goal here was to test hypothesis 1 by seeing whether there is the predicted effect of word complexity (a significant negative slope indicating that accuracy was lower for more complex items), hypothesis 2 by seeing whether there is the predicted effect of word and phrase frequency (a significant positive slope indicated that accuracy was greater for higher frequency words), and hypothesis 3 by seeing whether there is the predicted interaction between word complexity and our frequency predictors (a significant positive term for the interaction indicating that the effect of word complexity is lower for higher frequency words). This analysis procedure also allowed us to test the direction-agnostic prediction that word and phrase frequency would affect each other's effect on accuracy (a significant term in either direction for the interaction between word frequency and phrase frequency).

We tested these hypotheses using likelihood ratio tests in an iterative drop-one analysis. We first look at the effect of any interactions by building a full model and removing each interaction separately. Having removed any non-significant interactions but keeping all individual variables, we then perform a drop-one analysis on the remaining terms. We then remove any non-significant variables and repeat until only variables that explain significant variance, along with any component terms where interactions are included remain. Finally, in order to look at the contribution of the individual variables, we remove all remaining interactions where appropriate and perform a drop-one analysis on a model containing only the individual model terms.

When fixed effect terms were removed at each step, the random effects were left in for that term. For example, when comparing the full 2-way interaction model to the reduced comparison model the following models were used:

Full 2-way interaction model:

CORRECT ~ (1 + WORDFREQ*PHRASEFREQ + WORDFREQ*WCM + PHRASEFREQ*WCM |PARTICIPANT) + (1|POS) + (1|PHRASE) + WORDFREQ*PHRASEFREQ + WORDFREQ*WCM + PHRASEFREQ*WCM)

Reduced model evaluating WF*PF:

CORRECT ~ (1 + WORDFREQ*PHRASEFREQ + WORDFREQ*WCM + PHRASEFREQ*WCM |PARTICIPANT) + (1|POS) + (1|PHRASE) + WORDFREQ*WCM + PHRASEFREQ*WCM)

Here the WF*PF interaction is removed from the fixed effects, but random effects for the full 2-way interaction are still included. If a likelihood-ratio test showed that doing this did not harm fit, the fixed term is assumed to explain significant variance. Table 3-3 reports on all model comparisons performed.

Table 3-4 Fixed effects used in model comparisons

Term Evaluated	Larger Model	Reduced model	ChiSq	P value
WF*PF	WF*PF + WF*WCM + PF*WCM	WF*WCM + PF*WCM	1.0081	0.3154
WF*WCM	WF*PF + WF*WCM + PF*WCM	WF*PF + PF*WCM	10.152	0.001442 **
PF*WCM	WF*PF + WF*WCM + PF*WCM	WF*PF + WF*WCM	2.3221	0.1275
WF*WCM	PF + WF*WCM	WF + PF + WCM	12.634	0.0003787 ***
PF	PF + WF*WCM	WF*WCM	0.3063	0.58
WCM	WF + WCM	WF	0.015	0.9027
WF	WF + WCM	WCM	10.803	0.001013 **

Note: The random effect structure for a given set of fixed effects will differ depending on whether it is considered as a main or comparison model

As can be seen from Table 3-3, of the three interactions, only the interaction between word frequency and word complexity was found to explain significant unique variance. Of the three separate variables, only word frequency was found to explain significant unique variance.

These model comparisons suggest that the word frequency by word complexity interaction explains unique variance along with the word frequency term. We finally compared the WF*WCM model to a model including only word frequency (therefore removing WCM) and found that the interaction model (WF*WCM) was a significantly better fit than the comparison model ($\chi^2(2) = 12.933$, $p = .002$). Taken together we take these results as evidence that there is an effect of word frequency, and that the size of this effect is different for different word complexity values.

We report below the coefficients for a model containing only an interaction of word frequency and word complexity in Table 3-4. Random effects of participant, word position (adjective or noun) and item (phrase) on the intercept and a random effect of participant on all slopes were also included in the model.

Table 3-5 Coefficients for the Word Frequency*WCM interaction model

Predictor	Parameter estimates			
	Estimate	S.E	T value	P value
Intercept	1.15375	0.17744	6.502	<.001***
Word Freq	0.17515	0.05651	3.099	0.001939 **
WCM	-0.06114	0.05489	-1.114	0.265286
Word Freq*WCM	-0.17225	0.04747	-3.629	0.000285 ***

Table 3-4 shows that there is a significant main effect of word frequency. The positive estimate value indicates that there is a significant positive slope, where accuracy was greater for words with higher frequencies. There is also a significant interaction between word frequency and complexity. The negative estimate value indicates that there is a significant negative term, where the effect of word complexity increases for higher frequency words. Therefore, accuracy was greater for low complexity phrases when they contained high frequency words.

3.3.1 Second coding

20% of recordings were second coded by a researcher (see section 2.2.2.1 for method). The percentage of agreement and the Kappa score were calculated. Percentages of agreement for interrater reliability were 75.8 % and Kappa scores were 0.391. This Kappa score is on the edge of 'fair' to 'moderate' agreement (Cohen, 1960). To explore disagreements between coders, we also performed a robustness check as described in section 2.2.2.1. This was to check that the results that we report are not contingent on coder decisions about which there was not agreement.

To do this we first identified all specific phonemes in specific word contexts where there were disagreements that affected the correctness score given – phonemes which had been coded to be correct by one coder (the phoneme matched that of the target phoneme), and as a different phoneme by the other coder. When a target phoneme was coded as a different phoneme for at least two of the children, we allowed this different phoneme to be considered as an alternative form of the target. This meant that we allowed a match for both the main coder and the second coder, where that coder has said that the child produced the permitted alternative form. We then re-ran the relevant analyses with the alternative dataset to see how this new matching impacted results. The percentage of agreement and the Kappa score were also calculated for the alternative dataset. Percentages of agreement for interrater reliability were 87.8% and Kappa scores were 0.552. This Kappa score indicates 'moderate' agreement (Cohen, 1960). The results from the robustness check are described below.

Table 3-6 Fixed effects used in model comparisons

Term Evaluated	Larger Model	Reduced model	ChiSq	P value
WF*PF	WF*PF + WF*WCM + PF*WCM	WF*WCM + PF*WCM	3.3625	0.0667
WF*WCM	WF*PF + WF*WCM + PF*WCM	WF*PF + PF*WCM	5.6613	0.01734 *
PF*WCM	WF*PF + WF*WCM + PF*WCM	WF*PF + WF*WCM	14.601	0.0001329 ***
WF*WCM	PF + WF*WCM	WF + PF + WCM	10.628	0.001114 **
PF*WCM	WF + PF*WCM	WF + PF + WCM	16.413	<.001***
WF	WF + PF*WCM	PF*WCM	18.886	<.001 ***
PF	PF + WF*WCM	WF*WCM	0.2575	0.6118
WCM	WF + PF + WCM	WF + PF	0.2447	0.6209
PF	WF + PF + WCM	WF + WCM	0.0502	0.8227
WF	WF + PF + WCM	PF + WCM	20.325	<.001 ***

Note: The random effect structure for a given set of fixed effects will differ depending on whether it is considered as a main or comparison model

As can be seen from Table 3-5, of the three interactions, the interaction between word frequency and word complexity and the interaction between phrase frequency and word complexity were found to explain significant unique variance. However, of the three separate variables, just as in our main analysis, only word frequency was found to explain significant unique variance. While the inclusion of an interaction between phrase frequency and WCM improved fit relative to a model with only the individual terms in, neither of these variables were individually found to explain unique variance and we therefore remove the interaction term from consideration.

These model comparisons suggest that the word frequency by word complexity interaction explains unique variance along with the word frequency term. Taken together we take these results as evidence that there is an effect of word frequency, and that the size of this effect is different for different word complexity values. We report below the coefficients for a model containing only an interaction of word frequency and word complexity in Table 3-6. This secondary analysis confirms that our observed pattern of results is robust to coder disagreements.

Table 3-7 Coefficients for the Word Frequency*WCM model

Predictor	Parameter estimates			
	Estimate	S.E	T value	P value
Intercept	1.530978	0.139800	10.951	<.001 ***
Word Freq	0.261664	0.060284	4.341	<.001 ***
WCM	-0.005861	0.057928	-0.101	0.91941
Word Freq*WCM	-0.171519	0.053095	-3.230	0.00124 **

3.4 Interim discussion

The purpose of this experiment was to explore the effect of phrase frequency, word frequency and word complexity (and their interactions) on the accuracy of phoneme production in 3-year-olds. We are interested in whether frequency effects in children are attributable to phonological/motor chunking, as is widely thought to be the case in adults. To explore this, we built models which included main effects of phrase frequency, word frequency and complexity and their relevant interactions. This meant that we were able to explore whether word frequency, phrase frequency and complexity influenced the number of errors made by children. We were also able to explore whether there was an interaction between these effects.

The pattern of results was rather different from that which we anticipated. We predicted that there would be a main effect of word and phrase frequency, where the children's ability to repeat phonemes in phrases would be affected by the frequency of the phrases and their component words. The data were partly consistent with this prediction, in that we did find a main effect of word frequency, where children made fewer errors when producing phrases which contained high frequency words. However, no effect of phrase frequency was seen. We also predicted that there would be a main effect of word complexity, where children would be better at producing words which were less complex regardless of frequency. However, there was no such effect of complexity.

Finally, we predicted that the negative effect of WCM on children's ability to repeat phonemes contained in each word would be reduced when that word is high frequency or part of a high frequency phrase. So, children would be better at producing high complexity phrases when they were high in frequency or when they contained high frequency words. We anticipated that no effect of frequency would be seen for low complexity words as no chunking would be required for fluent production for such items. However, the interaction we observed was in the opposite direction. Children were better at producing low complexity words, but only when they were high in frequency. Rather than being fluent with the less complex words regardless of frequency, and displaying fluency with the more complex words only when practice had resulted in phonological/motor chunking, complexity had little impact (participant mean accuracy of 71.6% rather than 68.9% using the broad bands described above) for low frequency words. For high frequency words a greater effect of complexity (participant mean accuracy of 79.3% rather than 73.2%) was seen.

There are a couple of possible explanations for this pattern of data, and its deviation from our expectations. One is that the task is simply harder at the articulatory level than we thought and young children struggle with all words and phrases except where phonological/motor chunking has occurred due to high frequency. According to this account the effect of frequency would still be at the formulation level – the children can be assumed to have successfully retrieved the lemmas, including all phonological information, for the words produced just like adult speakers but struggle to produce them except where they have had extensive practice. Under this account the frequency effect could be attributable to phonological/motor chunking, but chunking that would only occur for easier forms at the age of three.

However, we cannot rule out an alternative explanation – that the word frequency effect is occurring during an earlier stage in production than predicted. This would mean that rather than retrieving full lemmas that they then have problems realising, the variance in children's ability to produce the different forms was due to differences in their knowledge, or in their ability to retrieve knowledge of, different forms. Specifically, it is possible that children have more robust knowledge of, or a more robust retrieval process for, high frequency words. To summarise, it might be that articulatory difficulty makes little difference if one does not have the knowledge of the lemmas that are required to begin the formulation process.

This analysis then is inconclusive with regard to the question of whether phonological/motor chunking is occurring. Fortunately, there is an additional analysis we can do to further address this question. Instead of looking at the accuracy of production we took those responses on which children were 100% correct and examined their fluency by looking at the duration of their productions. Since one possible explanation for the above findings is that children have incomplete knowledge of low frequency forms, by measuring the duration of only the correct productions, we set aside this possibility of incomplete knowledge. Children must have full conceptual knowledge of the word, and robust enough phonological representations to produce these phrases correctly. This means that we can explore differences in the speed of production for phrases, so that any effect of frequency, and or complexity, seems more likely to be due to the second, phonological realisation stage.

3.5 Duration data

To explore the effect of word frequency, phrase frequency and complexity on production in children, we next measured the duration of productions for phrases which were produced correctly, using the recordings from the previous experiment. Following the same logic as for experiment one as to what would constitute evidence for phonological/motor chunking, we tested the following hypotheses:

1) Children will be faster to repeat low complexity words.
2) Children will be faster to repeat high frequency words and phrases.
3) The effect of word complexity on children's speed of production will be reduced when that word is high frequency or part of a high frequency phrase.

3.6 Results

To address our research questions, we built a series of multilevel linear regression models. We started by building a model including phrase frequency, word frequency, word complexity and their relevant 2-way interactions as fixed effects. Phoneme count was also included as a covariate in all models in order to account for word length. Random effects of participant, part of phrase (adjective or noun) and item (phrase) on the intercept and a random effect of participant on all slopes were also included in all models.

Our goal here was to test hypothesis 1 by seeing whether there is the predicted effect of word complexity (a significant positive slope indicating that duration was greater for more complex items even once length in phonemes was taken into account), hypothesis 2 by seeing whether there is the predicted effect of word and phrase frequency (a significant negative slope indicating that productions were shorter for higher words and frequencies), and hypothesis 3 by seeing whether there is the predicted interaction between word complexity and our frequency predictors (a significant negative term for the interaction indicating that the effect of word complexity is lower for higher frequency words). This analysis procedure will also allow us to test the direction-agnostic prediction that word and phrase frequency will affect each other's effect on production duration (a significant term in either direction for the interaction between word frequency and phrase frequency).

The method for testing each model term was the same as for the error analysis. Starting with our full model, fixed effects were removed in an iterative drop-one procedure, while leaving the random slope for that term in the model (see section 2.3.3).

Table 3-8 Fixed effects used in model comparisons

Term Evaluated	Larger Model	Reduced model	ChiSq	P value
WF*PF	WF*PF + WF*WCM + PF*WCM	WF*WCM + PF*WCM	0.418	0.5179
WF*WCM	WF*PF + WF*WCM + PF*WCM	WF*PF + PF*WCM	1.5325	0.2157
PF*WCM	WF*PF + WF*WCM + PF*WCM	WF*PF + WF*WCM	1.9433	0.1633
WCM	WF + PF + WCM	WF + PF	0.2097	0.647
PF	WF + PF + WCM	WF + WCM	0.5551	0.4562
WF	WF + PF + WCM	PF + WCM	22.291	<.001 ***

Note: The random effect structure for a given set of fixed effects will differ depending on whether it is considered as a main or comparison model

The model comparisons performed can be seen in Table 3-7. None of the interactions were found to explain unique variance. Of the individual predictors, only word frequency was found to explain unique variance.

We take these results as evidence that there is an effect of word frequency. We report the coefficients for a model containing word frequency and phoneme count in Table 3-8. Random effects of participant, word position (adjective or noun) and item (phrase) on the intercept and a random effect of participant on fixed effects were also included in the model.

Table 3-9 Coefficients for the Word Frequency model

Predictor	Parameter estimates			
	Estimate	S.E	T value	P value
Intercept	0.476860	0.071418	6.677	0.0794
Word Freq	-0.049404	0.009575	-5.160	<.001***
Phoneme Count	0.045408	0.007792	5.828	<.001***

Table 3-8 shows that there is a significant main effect of word frequency. The negative estimate value indicates that there is a significant negative slope, where duration was shorter for words with higher frequencies.

3.6.1 Second coding

To check that the results we report above are not contingent on coder decisions about which there was not agreement in the error data, we re-ran the relevant analyses with the alternative dataset created above (see section 3.3.1). To do this, we used the phrases which were produced correctly based on the alternative dataset.

Table 3-10 Fixed effects used in model comparisons

Term Evaluated	Larger Model	Reduced model	ChiSq	P value
WF*PF	WF*PF + WF*WCM + PF*WCM	WF*WCM + PF*WCM	0.6965	0.404
WF*WCM	WF*PF + WF*WCM + PF*WCM	WF*PF + PF*WCM	2.4765	0.1156
PF*WCM	WF*PF + WF*WCM + PF*WCM	WF*PF + WF*WCM	5.6711	0.01725 *
WF	WF + PF*WCM	PF*WCM	18.154	<.001 ***
PF*WCM	WF + PF*WCM	WF	3.4147	0.06462
WCM	WF + PF + WCM	WF + PF	0.7338	0.3917
PF	WF + PF + WCM	WF + WCM	0.6258	0.4289
WF	WF + PF + WCM	PF + WCM	18.887	<.001 ***

Note: The random effect structure for a given set of fixed effects will differ depending on whether it is considered as a main or comparison model

As can be seen from Table 3-9, of the three interactions, the interaction between phrase frequency and word complexity is found to explain significant unique variance. However, of the three separate variables, only word frequency was found to explain significant unique variance. While the inclusion of an interaction between phrase frequency and WCM improved fit relative to a model with only the individual terms in, neither of these variables were individually found to explain unique variance and we therefore remove the interaction term from consideration.

Therefore, we take these results as evidence that there is an effect of word frequency only. We report below the coefficients for a model containing only an interaction of word frequency in Table 3-10. This secondary analysis confirms that our observed pattern of results is robust to coder disagreements.

Table 3-11 Coefficients for the Word Frequency model

Predictor	Parameter estimates			
	Estimate	S.E	T value	P value
Intercept	0.479982	0.064733	7.415	0.067
Word Freq	-0.042376	0.009069	-4.673	<.001 ***
Phoneme Count	0.039791	0.007045	5.648	<.001 ***

3.7 Interim discussion

The purpose of this experiment was to explore the effect of phrase frequency, word frequency, word complexity (and their interactions) on speed of production in 3-year-olds.

We are interested in whether frequency effects in children are attributable to phonological/motor chunking when we look only at those items for which all phonemes were correctly produced (thereby reducing the impact of conceptual or lemma knowledge). To explore this, we built models which included main effects of phrase frequency, word frequency and complexity and their relevant interactions. Phoneme count was also included as a covariate.

We predicted that there would be a main effect of word complexity, where children would be faster in producing words that were less complex. However, there was no main effect of complexity. We also predicted that there would be a main effect of word and phrase frequency, where the children's ability to repeat phonemes in phrases would be affected by the frequency of the phrases and their component words. The data were partly consistent with this prediction, in that we did find a main effect of word frequency, where speakers were faster to produce phrases which contained high frequency words. However, no effect of phrase frequency was seen.

Finally, we predicted that the effect of WCM on children's speed of production would be reduced when that word was high frequency or part of a high frequency phrase. So, children would be better at producing high complexity phrases when they were high in frequency or when they contained high frequency words. No significant interaction was seen. This finding suggests that once we ensure that children have robust conceptual and phonological knowledge of the target phrase, they will start to show an advantage for producing high frequency words regardless of complexity. This pattern of results is not inconsistent with the claim that phonological/motor chunking is occurring but is not as strong evidence for it as if we had seen an interaction between frequency and complexity.

3.8 Discussion

The purpose of this experiment was to explore the effect of phrase frequency, word frequency, word complexity (and their interactions) on speech production in 3-year-olds. We are interested in whether frequency effects in children are attributable to phonological/motor chunking, as is generally accepted to be the case in adults. To explore this, we built models which included main effects of phrase frequency, word frequency and complexity (and their relevant interactions).

The error data showed an effect of word frequency and plausibly an effect of word complexity but only for high frequency words. This finding goes against our original prediction that children would be fluent with the less complex words regardless of frequency and would be more fluent for complex words only when practice had resulted in phonological/motor chunking.

We proposed 2 possible explanations for this finding. One possibility is that the task was simply harder at the articulatory level than we thought and young children are only able to produce the words when they are high frequency. According to this account, the effect of frequency would still be at the formulation level. The children can be assumed to have successfully retrieved the lemmas but struggle to produce them except where they were very articulatorily straight-forward and they had had extensive practice that had given rise to phonological/motor chunking.

A second possibility, however, was that the word frequency effect was occurring during an earlier stage in production than we originally assumed would be the case. Under this account, no phonological/motoric chunking would be involved. Our results show that children were best at producing phrases when they contained high frequency words that were low in complexity. It is possible that children might have more robust knowledge of, or a more robust retrieval process for, high frequency words. Once these high frequency words have been retrieved, speakers find it more difficult to produce the high complexity words in comparison to the low complexity words. Furthermore, our results show that children did not differ in their error rate for words of varying complexity when they were low in frequency. This might be because children do not (at this age) have strong representations for the lemmas of words which are not frequent.

We next performed an analysis of the duration of the children's productions, including only those trials on which they produced the correct sequence of phonemes. The aim of this analysis was to explore the fluency of production for phrases for which children have complete knowledge. We suggest that children must have full conceptual knowledge of the phrase, and robust enough phonological representations to produce these phrases correctly (we assume that children were able to retrieve and activate the correct conceptual information and lemmas for these target phrases before activation of the correct phonological representations followed by accurate motor movements during articulation). However, it could be argued that the children's knowledge of these phrase is partial in nature. It is possible that children do not produce these words/phrase correctly 100% of the time in natural speech.

We found that children produced phrases more quickly when they contained words which were high in frequency. There was no effect of complexity or interaction between frequency and complexity. Nonetheless, the results provide tentative support for the claim that phonological/motoric chunking is occurring.

To further explore these explanations, it will be necessary to begin to look at how the pattern changes with development. In the next chapter we will repeat this same experimental paradigm but with slightly older children.

4 Chapter 4 - Assessing the effect of word frequency, phrase frequency and phonological complexity on fluency in 4-year-olds

4.1 Introduction

Both adults and children have been shown to be more fluent (are faster and/or make fewer errors) when producing words and phrases which are more frequent compared to words and phrases which are less frequent. Sosa and Stoel-Gammon (2012) found that children are more variable in their production of low frequency words, while both children and adults improve in their production of novel non-words after practice (Sasisekaran, Smith, Sadagopan & Weber-Fox, 2010). As for phrases, children make fewer errors and take less time to produce frequent phrases compared to infrequent phrases (Bannard & Matthews, 2008), while adults produce phrases more quickly when they are frequent (Arnon & Cohen-Priva, 2013).

One interpretation as to why adults are better at producing words which are more frequent in the language is phonological/motor chunking. Here, the production system is said to store frequent phoneme sequences as a single ‘item’ – or chunk. Guenther (2016) proposes that during the formulation stage, working memory stores frequent chunks to reduce phonological working memory load. This means that speakers can access this stored information more quickly. These chunks are then released to the motor system, where the motor routines for these sequences are also more efficient due to repeated production practice. Targets which are not stored as chunks take longer to produce. In this case, the production system first loads the target sequence, then the phonological content is selected based on activity gradient. Next, phonological items are selected before the best motor programs are chosen. Finally, the relevant phonemes are activated in sequence.

It is not clear whether frequency effects observed in children can be attributed to phonological/motor chunking in the same way. In the previous chapter we aimed to explore whether frequency effects in 3-year-old children are due to phonological/motor chunking. We were also interested in how phrase frequency would affect fluency in 3-year-olds since it is not clear at which stage of production phrase frequency effects occur. To do this we used an elicited imitation task where we crossed word frequency, phrase frequency and complexity over 24 adjective-noun combinations.

We reasoned that if, like adults, children do engage in phonological/motor chunking, children would find words that are difficult to articulate easier to say after practice. Therefore, we predicted that for complex words, children would make fewer errors when producing phrases that were high in frequency or contained high frequency words, compared to complex phrases which were low in frequency or contained low frequency words. We also predicted that children would make very few errors when producing low complexity words – regardless of frequency. This is based on the assumption that since low complexity words are inherently easy to produce, children would not need practice for these items. However, the observed pattern of results was different. Children were better at producing phrases when they were low in complexity, but only when the words were high in frequency, whereas when words were low in frequency, they were seemingly unaffected by complexity.

We next proceeded to measure the duration of phrases which were produced correctly. We found that children were better at producing phrases when they contained high frequency words. However, we found no effect of complexity and no interaction between frequency and complexity. We take this as evidence that since children must have full conceptual knowledge and robust enough phonological representations to produce the analysed phrases correctly, one interpretation of these results is that once the children are able to retrieve the lemma, phonological/motor chunking takes place regardless of phonological complexity.

In this chapter, we look at how the pattern of results might change with development. If the results from experiment one do reflect the fact that the task was simply more difficult for 3-year-old children than anticipated, we might expect older children to be less affected by this articulatory complexity issue. In this case older children would, as we originally predicted for the 3-year-olds, make few errors when producing low complexity words – regardless of frequency.

We tested older children using the same methodology as for experiment 1. Children aged 3;9 - 4;4 were asked to repeat 24 adjective-noun phrases which were manipulated for phrase frequency, word frequency and word complexity. Each phoneme in each target sequence was coded as accurately produced or not accurately produced. As in experiment 1, we aim to explore whether frequency effects in children are attributable to phonological/motor chunking. We also measured the duration of correct productions in the second part of this experiment (see section 2.2.3).

Therefore, our hypotheses were the same as for experiment 1:

1. Children will be more accurate in producing phonemes when repeating low complexity words
2. Children's ability to repeat phonemes in phrases will be affected by the frequency of the phrases and their component words
3. The negative effect of WCM on children's ability to repeat phonemes contained in each word will be reduced when that word is high frequency or part of a high frequency phrase

We also again test the secondary prediction that phrase and word frequency will affect (promote or suppress) each other's effect on accuracy.

4.2 Method

This study was preregistered on the Open Science Framework website (https://osf.io/3pkbx).

4.2.1 Participants

49 monolingual English speaking 3- and 4-year-olds who were aged between 3;9 and 4;4 (mean age: 4;0) were recruited and included in analyses. Parental consent was given, and

this study was approved by the ethics board. All children were tested in preschools and nurseries in Bedfordshire and Buckinghamshire. 4 additional children were tested but were excluded (exclusion criteria in OSF). 3 children refused to take part in the experiment and 1 child had a diagnosed language disorder.

A sample size of 65 was determined in advance (and preregistered) based on the same power analysis as experiment one. However due to the 2020 pandemic situation and consequent school closures, data collection ceased after 49 children had been tested.

4.2.2 Design and Materials

The design and materials were the same as for experiment 1 (see section 3.2.2 and 3.2.3)

4.2.3 Procedure

The procedure was the same as for experiment 1 (see section 3.2.4).

4.3 Results

To address our research questions, we built a series of multilevel logistic regression models. We started by building a model including phrase frequency, word frequency, word complexity and their relevant 2-way interactions as fixed effects. Random effects of participant, part of phrase (adjective or noun) and item (phrase) on the intercept and a random effect of participant on fixed effects were also included in the model.

Our goal here was to test hypothesis 1 by seeing whether there is the predicted effect of word complexity (a significant negative slope indicating that accuracy was lower for more complex items), hypothesis 2 by seeing whether there is the predicted effect of word and phrase frequency (a significant positive slope indicated that accuracy was greater for higher frequency words), and hypothesis 3 by seeing whether there is the predicted interaction between word complexity and our frequency predictors (a significant positive term for the interaction indicating that the effect of word complexity is lower for higher frequency words). This analysis procedure also allowed us to test the direction-agnostic prediction that word and phrase frequency would affect each other's effect on accuracy (a significant term in either direction for the interaction between word frequency and phrase frequency).

We tested these hypotheses using likelihood ratio tests in an iterative drop-one analysis. We first look at the effect of any interactions by building a full model and removing each interaction separately. Having removed any non-significant interactions but keeping all individual variables, we then perform a drop-one analysis on the remaining terms. We then remove any non-significant variables and repeat until only variables that explain significant variance, along with any component terms where interactions are included. Finally, to look at the contribution of the individual variables, we remove all remaining interactions where appropriate and perform a drop-analysis on a model containing only the individual model terms.

Table 4-1 Fixed effects used in model comparisons

Term Evaluated	Larger Model	Reduced model	ChiSq	P value
WF*PF	WF*PF + WF*WCM + PF*WCM	WF*WCM + PF*WCM	0	1
WF*WCM	WF*PF + WF*WCM + PF*WCM	WF*PF + PF*WCM	12.141	0.0004933 ***
PF*WCM	WF*PF + WF*WCM + PF*WCM	WF*PF + WF*WCM	0.3431	0.558
WF*WCM	PF + WF*WCM	WF + PF + WCM	11.921	0.0005549 ***
PF	PF + WF*WCM	WF*WCM	0.0808	0.7762
WF	WF + WCM	WCM	4.3988	0.03596 *
WCM	WF + WCM	WF	0.0617	0.8039

Note: The random effect structure for a given set of fixed effects will differ depending on whether it is considered as a larger or reduced model

As can be seen from Table 4-1, of the three interactions, only the interaction between word frequency and word complexity was found to explain significant unique variance. Of the three separate variables, only word frequency was found to explain significant unique variance.

These model comparisons suggest that the word frequency by word complexity interaction explains unique variance along with the word frequency term. We finally compared the

WF*WCM model to a model including only word frequency (therefore removing WCM) and found that the interaction model (WF*WCM) was a significantly better fit than the comparison model ($\chi2(2) = 12.14$, $p = .002$). Taken together we take these results as evidence that there is an effect of word frequency, and that the size of this effect is different for different word complexity values.

In Table 4-2 we report the coefficients for a model containing an interaction of word frequency and word complexity. Random effects of participant, word position (adjective or noun) and item (phrase) on the intercept and a random effect of participant on the interaction were also included in the model.

Table 4-2 Coefficients for the Word Frequency*WCM interaction model

Predictor	Parameter estimates		Wald's test	
	Logs-odds	S.E	Z	p
Intercept	1.70580	0.22321	7.642	<.001 ***
Word Freq	0.10820	0.06647	1.628	0.103538
WCM	-0.01944	0.07100	-0.274	0.784276
Word Freq*WCM	-0.25009	0.07255	-3.447	0.000567 ***

Table 4-2 shows that there is a significant main effect of word frequency. The positive estimate value indicates that there is a significant positive slope, where accuracy increased for words with higher frequencies. There is also a significant interaction between word frequency and complexity. The negative estimate value indicates that there is a significant negative term, where the effect of word complexity increases for higher frequency words. Therefore, accuracy was greater for low complexity phrases when they contained high frequency words.

4.3.1 Second coding

20% of recordings were second coded by a researcher (see section 2.2.1 for method). The percentage of agreement and the Kappa score were calculated. Percentages of agreement for interrater reliability were 81.7% and Kappa scores were 0.477. This Kappa score indicates ‘moderate’ agreement (Cohen, 1960). To explore disagreements between coders, we also performed a robustness check as described in section 2.2.2.1. This was to check that the results that we report are not contingent on coder decisions about which there was not agreement. We re-ran the relevant analyses with the alternative dataset to see how this new matching impacted results. The percentage of agreement and the Kappa score were also calculated for the alternative dataset. Percentages of agreement for interrater reliability were 91.4% and Kappa scores were 0.647. This Kappa score indicates ‘substantial’ agreement (Cohen, 1960). The results from the robustness check are described below.

Table 4-3 Fixed effects used in model comparisons

Term Evaluated	Larger Model	Reduced model	ChiSq	P value
WF*PF	WF*PF + WF*WCM + PF*WCM	WF*WCM + PF*WCM	0	1
WF*WCM	WF*PF + WF*WCM + PF*WCM	WF*PF + PF*WCM	7.0155	0.008081 **
PF*WCM	WF*PF + WF*WCM + PF*WCM	WF*PF + WF*WCM	3.7973	0.05134
WF*WCM	PF + WF*WCM	WF + PF + WCM	2.8981	0.08868
PF	PF + WF*WCM	WF*WCM	0.8488	0.3569
WCM	WF + WCM	WF	1.1255	0.2887
WF	WF + WCM	WCM	4.5159	0.03358 *

Note: The random effect structure for a given set of fixed effects will differ depending on whether it is considered as a main or comparison model

As can be seen from Table 4-3, of the three interactions, the interaction between word frequency and word complexity was found to explain significant unique variance. However, of the three separate variables, only word frequency was found to explain significant unique variance. These model comparisons suggest that the word frequency by word complexity interaction explains unique variance along with the word frequency term. We report below the coefficients for a model containing only an interaction of word frequency and word complexity in Table 4-4.

Table 4-4 Coefficients for the Word Frequency*WCM interaction model

Predictor	Parameter estimates		Wald’s test	
	Estimate	S.E	T value	P value
Intercept	1.96675	0.18724	10.504	<.001 ***
Word Freq	0.12750	0.07134	1.787	0.0739
WCM	0.07320	0.07329	0.999	0.3179
Word Freq*WCM	-0.13499	0.07646	-1.765	0.0775

The results from the second coding analysis show that we have a main effect of word frequency, where children make fewer errors when producing words which are more frequent. However, Table 4-4 shows that the interaction effect is marginal in this secondary analysis. While we take the two analyses together as evidence that this interaction is worthy of discussion, it is not completely robust to annotator disagreements here and needs to be confirmed in further studies.

4.4 Interim discussion

The purpose of this experiment was to explore the effect of phrase frequency, word frequency and word complexity (and their interactions) on the accuracy of phoneme production in older children. We predicted that there would be a main effect of word complexity, where children would be better at producing words that were less complex. However, there was no main effect of complexity. We also predicted that there would be a main effect of word and phrase frequency, where the children's ability to repeat phonemes in phrases would be affected by the frequency of the phrases and their component words. We observed an effect of word frequency but no effect of phrase frequency.

Finally, and perhaps most critically for our research questions, we predicted that the negative effect of WCM on children's ability to repeat phonemes contained in each word would be reduced when that word is high frequency or part of a high frequency phrase. So, children would be better at producing high complexity phrases when they were high in frequency, or when the phrase contained high frequency words. However, like for the 3-year-old data, we found the opposite prediction to be true. Children were better at producing phrases when they contained high frequency words, but only when they were low in complexity. Rather than being fluent with the less complex words regardless of frequency, and displaying fluency with the more complex words only when practice had resulted in phonological/motor chunking, the children struggled with all complex items, and were only fluent with the less complex items when they were frequent.

Since these results mirror those in the 3-year-old data there are the same possible explanations for this pattern of data. It is still possible that the task is harder at the articulatory level than we expected for children at this age. These children could be showing phonological/motor chunking but only for the least complex items. According to this account, the effect of frequency would still be at the formulation level – the children can be assumed to have successfully retrieved the lemmas, including all phonological information, for the words produced, just like adult speakers, but struggle to produce them except where they were very articulatorily straight-forward and they had had extensive practice. As these children are older, this interpretation would seem slightly less likely but still plausible.

Again, it is also possible that the word frequency effect is occurring during an earlier stage in production where the variance in children's ability to produce the different forms was due to differences in their knowledge, or in their ability to retrieve knowledge of, different forms. Therefore, children might have more robust knowledge of, or a more robust retrieval process for high frequency words – where articulatory difficulty makes little difference.

To further tease apart these explanations, we next explored the effect of phrase frequency, word frequency and complexity on the duration of correctly produced phrases. To do this, we used the recordings from the above experiment and measured the duration of phrases which were produced correctly, like in experiment 1. If the word frequency effect in the above experiment is occurring during an earlier stage in production, once we remove the problem of whether children have a robust retrieval process for the target phrase, we would expect children to be faster when producing high complexity phrases when they were high in frequency or when they contained high frequency words, as a result of motor learning.

4.5 Duration data

To explore the effect of word frequency, phrase frequency and complexity on production in children, we measured the duration of productions for words which were produced correctly, using the recordings from the error data.

We aimed to test the same hypotheses as in experiment 1:

1) Children will take less time to repeat low complexity words.
2) Children will take less time to repeat high frequency words and phrases.
3) The effect of word complexity on children's speed of production will be reduced when that word is high frequency or part of a high frequency phrase.

We also again test the secondary prediction that phrase and word frequency will affect (promote or suppress) each other's effect on accuracy.

4.6 Results

To address our research questions, we built a series of multilevel linear regression models. We started by building a model including phrase frequency, word frequency, word complexity and their relevant 2-way interactions as fixed effects. Phoneme count was also included as a covariate. Random effects of participant, part of phrase (adjective or noun) and item (phrase) on the intercept and a random effect of participant on fixed effects were also included in the model.

Our goal here was to test hypothesis 1 by seeing whether there is the predicted effect of word complexity (a significant positive slope indicating that duration was greater for more complex items even once length in phonemes was taken into account), hypothesis 2 by seeing whether there is the predicted effect of word and phrase frequency
(a significant negative slope indicating that productions were shorter for higher words and frequencies), and hypothesis 3 by seeing whether there is the predicted interaction

between word complexity and our frequency predictors (a significant negative term for the interaction indicating that the effect of word complexity is lower for higher frequency words). This analysis procedure will also allow us to test the direction-agnostic prediction that word and phrase frequency will affect each other's effect on production duration (a significant term in either direction for the interaction between word frequency and phrase frequency).

Again, we tested these hypotheses using likelihood ratio tests in an iterative drop-one analysis. We first look at the effect of any interactions by building a full model and removing each interaction separately. Having removed any non-significant interactions but keeping all individual variables, we then perform a drop-one analysis on the remaining terms. We then remove any non-significant variables and repeat until only variables that explain significant variance, along with any component terms where interactions are included. Finally, to look at the contribution of the individual variables, we remove all remaining interactions where appropriate and perform a drop-one analysis on a model containing only the individual model terms.

Table 4-5 Fixed effects used in model comparisons

Term Evaluated	Larger Model	Reduced model	ChiSq	P value
WF*PF	WF*PF + WF*WCM + PF*WCM	WF*WCM + PF*WCM	0.4424	0.506
WF*WCM	WF*PF + WF*WCM + PF*WCM	WF*PF + PF*WCM	6.9992	0.008155 **
PF*WCM	WF*PF + WF*WCM + PF*WCM	WF*PF + WF*WCM	7.3845	0.006579 **
WF*WCM	WF*WCM + PF*WCM	WF + PF*WCM	10.922	0.0009504 ***
PF*WCM	WF*WCM + PF*WCM	WF*WCM + PF	7.9723	0.00475 **
WCM	WF + WCM	WF	1.4697	0.2254
WF	WF + WCM	WCM	32.193	<.001 ***
PF	WF + PF	WF	2.9798	0.08431

Note: The random effect structure for a given set of fixed effects will differ depending on whether it is considered as a main or comparison model

As can be seen from Table 4-5, of the three interactions, both the interaction between word frequency and word complexity and that between phrase frequency and word complexity were found to explain significant unique variance. Of the three separate variables, only word frequency was found to explain significant unique variance.

These model comparisons suggest that the word frequency by word complexity interaction explains unique variance along with the word frequency term. We finally compared the WF*WCM model to a model including only word frequency (therefore removing WCM) and found that the interaction model (WF*WCM) was a significantly better fit than the comparison model ($\chi^2(2) = 8.72$, $p = 0.013$). While the inclusion of an interaction between phrase frequency and WCM improved fit relative to a model with only the individual terms in, neither of these variables were individually found to explain unique variance and we therefore remove the interaction term from consideration. Taken together, we take these results as evidence that there is an effect of word frequency, but one that varies depending on the level of complexity

We report below the coefficients for a model containing only an interaction of word frequency and word complexity in Table 4-6. Random effects of participant, word position (adjective or noun) and item (phrase) on the intercept and a random effect of participant on all slopes were also included in the model.

Table 4-6 Coefficients for the Word Frequency*WCM interaction model

Predictor	Parameter estimates			
	Estimate	S.E	T value	P value
Intercept	0.534036	0.070277	7.599	0.0668
Word Freq	-0.053718	0.008961	-5.995	<.001 ***
WCM	0.022526	0.013643	1.651	0.1022
Phoneme Count	0.023421	0.013677	1.712	0.0925
Word Freq*WCM	-0.018470	0.008354	-2.211	0.0283 *

Table 4-6 shows that there is a significant main effect of word frequency. The negative estimate value indicates that there is a significant negative slope, where duration was shorter for words with higher frequencies. There is also a significant word frequency by complexity interaction. The negative estimate value indicates that there is a significant negative term for the interaction, where the effect of word complexity is lower for higher frequency words. So, duration was shorter for high complexity words which were also high in word frequency.

4.6.1 Second coding

To check that the results we report above are not contingent on coder decisions about which there was not agreement in the error data, we re-ran the relevant analyses with the alternative dataset created above (see section 4.3.1). To do this, we used the phrases which were produced correctly based on the alternative dataset.

Table 4-7 Fixed effects used in model comparisons

Term Evaluated	Larger Model	Reduced model	ChiSq	P value
WF*PF	WF*PF + WF*WCM + PF*WCM	WF*WCM + PF*WCM	0.0016	0.9683
WF*WCM	WF*PF + WF*WCM + PF*WCM	WF*PF + PF*WCM	5.8206	0.01584 *
PF*WCM	WF*PF + WF*WCM + PF*WCM	WF*PF + WF*WCM	10.51	0.001187 **
WF*WCM	WF*WCM + PF*WCM	WF + PF*WCM	6.0903	0.01359 *
PF*WCM	WF*WCM + PF*WCM	WF*WCM + PF	12.917	0.0003256 ***
WCM	WF + WCM	WF	2.9453	0.08613
WF	WF + WCM	WCM	37.073	<.001 ***
WCM	PF + WCM	PF	7.7992	0.005227 **
PF	PF + WCM	WCM	0.262	0.6087

Note: The random effect structure for a given set of fixed effects will differ depending on whether it is considered as a main or comparison model

As can be seen from Table 4-7, of the three interactions, both the interaction between word frequency and word complexity and that between phrase frequency and word complexity were found to explain significant unique variance. Of the three separate variables, word frequency and complexity were found to explain significant unique variance.

We take these results as evidence that there is an interaction between word frequency and complexity. In addition to this, there is an effect of complexity for certain levels of phrase frequency. We report below the coefficients for a model containing an interaction of word frequency and word complexity and an interaction of phrase frequency and word complexity in Table 4-8. Random effects of participant, word position (adjective or noun) and item (phrase) on the intercept and a random effect of participant on all slopes were also included in the model.

Table 4-8 Coefficients for the Word Frequency*WCM interaction model

Predictor	Parameter estimates			
	Estimate	S.E	T value	P value
Intercept	0.431472	0.067031	7.897	0.064084

Word Freq	-0.054883	0.008538	-6.428	<.001 ***
WCM	0.014617	0.012060	1.212	0.226863
Phrase Freq	0.020451	0.012846	1.592	0.129101
Phoneme Count	0.022333	0.010611	2.105	0.038236 *
Word Freq*WCM	-0.020083	0.008106	-2.478	0.014249 *
WCM*Phrase Freq	0.030632	0.008505	3.602	0.000404 ***

We take the two analyses together as evidence for a main effect of word frequency and for a word frequency by complexity interaction. However, the significant phrase frequency by complexity interaction indicates that these results are not completely robust to annotator disagreements. This needs to be confirmed in further studies.

4.7 Interim discussion

The purpose of this experiment was to explore the effect of phrase frequency, word frequency, word complexity (and their interactions) on production duration in older children. We predicted that there would be a main effect of word complexity, where children would take less time to produce words that were less complex. We found no main effect of complexity. We also predicted that there would be a main effect of word and or phrase frequency, where children would take less time to repeat high frequency words and phrases. There was no main effect of phrase frequency. We did find a main effect of word frequency where children produced phrases containing high frequency words more quickly in comparison to phrases containing low frequency words.

Finally, we predicted that the effect of word complexity on children's speed of production will be reduced when that word is high frequency or part of a high frequency phrase. So, children would be better at producing high complexity phrases when they were high in frequency or when they contained high frequency words. The data were consistent with the second part of this prediction.

These results are in line with our original prediction (which we did not find support for in the accuracy analysis). It appears then that once we remove the retrieval issue, looking only at phrases where children have robust lemmas, children are better at producing phrases which are high in word frequency and complexity. We tentatively interpret this as evidence that frequency is driving phonological/motor chunking which leads to a short production duration.

4.8 Combined age data

In the previous experiments, we found that both 3- and 4-year-old children made fewer errors for phrases containing high frequency words which were also low in complexity. However, we would expect older children to have more extensive knowledge of the target phrases given that they have had more exposure to adult speech. As well as this, older children are likely to have had more production practice. Therefore, we would expect older children to be more fluent in their productions than younger children. The results from the duration data tentatively support this idea, where older children show stronger evidence of phonological/motor chunking than younger children. The question we are asking here then is whether a combined analysis supports the idea that there is developmental change.

The following aims to explore the effect of age on fluency using the data from the previous experiments. We combined the 3-year-old and 4-year-old data and measured the effect of age in months with the previous significant fixed effects. One participant was excluded from the analysis because we did not have a record of their age in months. We predict that older children will be better at the task in general and so expect that they will make fewer errors and take less time to produce phrases overall. We also anticipate that there will be an interaction between age, word frequency and word complexity.

4.9 Combined error data

To explore the effect of age, word frequency and complexity on production in children, we combined the error data from the 3- and 4-year-old samples. Age was included in the models as a continuous variable using months.

We aim to test the following hypotheses:

1. Older children will be more accurate when producing phonemes in comparison to younger children
2. The negative effect of WCM on children’s ability to repeat phonemes contained in each word will be reduced for older children in comparison to younger children
3. The positive effect of word frequency on children’s ability to repeat phonemes contained in each word will be reduced for younger children
4. The negative effect of WCM on children’s ability to repeat phonemes contained in each word will be reduced when that word is high frequency or part of a high frequency phrase

4.10 Results

To address our research questions, we built a series of multilevel logistic regression models. We started by building a model including word frequency, word complexity, age in months and their relevant 2-way interactions as fixed effects. Random effects of participant, part of phrase (adjective or noun) and item (phrase) on the intercept and a random effect of participant on fixed effects were also included in the model.

Our goal here was to test hypothesis 1 by seeing whether there is the predicted effect of age (a significant positive slope would indicate that accuracy was higher for older children), hypothesis 2 by seeing whether there is the predicted interaction between age and word complexity (a significant positive term would indicate that older children had greater accuracy for more complex words), hypothesis 3 by seeing whether there is the predicted interaction between age and word frequency (a significant positive term would indicate that older children had greater accuracy for more frequent words), and hypothesis 4 by seeing whether there is the predicted interaction between word complexity and word frequency (a significant positive term for the interaction would indicate that the effect of word complexity is lower for higher frequency words). We tested these hypotheses using likelihood ratio tests in an iterative drop-one analysis, as in the previous experiments.

Table 4-9 Fixed effects used in model comparisons

Term Evaluated	Larger Model	Reduced model	ChiSq	P value
WF*WCM	WF*WCM + WF*AGE + WCM*AGE	WF*AGE + WCM*AGE	22.255	<.001 ***
WF*AGE	WF*WCM + WF*AGE + WCM*AGE	WF*WCM + WCM*AGE	1.2748	0.2589
WCM*AGE	WF*WCM + WF*AGE + WCM*AGE	WF*WCM + WF*AGE	4.1492	0.04165 *
WF*WCM	WF*WCM + WCM*AGE	WCM*AGE	40.348	<.001 ***
WCM*AGE	WF*WCM + WCM*AGE	WF*WCM	26.745	<.001 ***
WCM	WF + WCM + AGE	WF + AGE	0.4332	0.5104
WF	WF + WCM + AGE	WCM + AGE	15.391	<.001 ***
AGE	WF + WCM + AGE	WF + WCM	22.794	<.001 ***

Note: The random effect structure for a given set of fixed effects will differ depending on whether it is considered as a main or comparison model

As can be seen from Table 4-9, of the three interactions, both the interaction between word frequency and word complexity and that between word complexity and age were found to explain significant unique variance. Of the three separate variables, both word frequency and age were found to explain significant unique variance.

These model comparisons firstly suggest that the word frequency by complexity interaction explains unique variance along with the word frequency term. We compared the WF*WCM model to a model including only word frequency (therefore removing WCM) and found that the interaction model (WF*WCM) was a significantly better fit than the comparison model ($\chi2(2)$ =23.25, p <.001). The model comparisons also suggest that the age by complexity interaction explains unique variance along with the age term. We compared the AGE*WCM model to a model including only age (therefore removing WCM) and found that the interaction model (AGE*WCM) was not a significantly better fit than the comparison model ($\chi2(2)$ = 3,66 p =.161). Taken together, we take these results as evidence that there is an effect of age, and an effect of word frequency, but one that varies depending on the levels of complexity.

We report the coefficients for a model containing an interaction of word frequency and word complexity, and main effect of age in Table 4-10. Random effects of participant, word position (adjective or noun) and item (phrase) on the intercept and a random effect of participant on all slopes (excluding age) were also included in the model.

Table 4-10 Coefficients for the Word Frequency*WCM (complexity) + Age model

Predictor	Parameter estimates			
	Log-odds	S.E	z value	p
Intercept	1.39274	0.17702	7.868	<.001 ***
Word Freq	0.16316	0.04714	3.461	0.000538 ***
WCM	-0.02875	0.04732	-0.608	0.543493
Age	0.29285	0.05775	5.071	<.001 ***
Word Freq*WCM	-0.20004	0.04215	-4.746	<.001 ***

Table 4-10 shows that there is a significant main effect of word frequency. The positive estimate value indicates a positive slope where accuracy was higher for words higher in frequency. There is also a significant main effect of age. The positive estimate value indicates a positive slope, where accuracy was higher for older children. There was also a significant word frequency by complexity interaction. The negative estimate value indicates that there is a significant negative term, where the effect of word complexity increases for higher frequency words. Therefore, accuracy was greater for low complexity phrases when they contained high frequency words.

4.10.1 Second coding

To check that the results we report above are not contingent on coder decisions about which there was not agreement in the error data, we re-ran the relevant analyses with the combined 3- and 4-year-old alternative datasets created above (see section 3.3.1 and 4.3.1). Age in months was added as a continuous variable and analyses were run on this combined alternative dataset.

Table 4-11 Fixed effects used in model comparisons

Term Evaluated	Larger Model	Reduced model	ChiSq	P value
WF*WCM	WF*WCM + WF*AGE + WCM*AGE	WF*AGE + WCM*AGE	11.612	0.0006554 ***
WF*AGE	WF*WCM + WF*AGE + WCM*AGE	WF*WCM + WCM*AGE	6.5485	0.0105 *
WCM*AGE	WF*WCM + WF*AGE + WCM*AGE	WF*WCM + WF*AGE	3.7577	0.05256
WF*WCM	WF*WCM + WCM*AGE	WCM*AGE	39.822	<.001 ***
WCM*AGE	WF*WCM + WCM*AGE	WF*WCM	16.573	0.0002519 ***
WF*WCM	WF*WCM + WF*AGE	WF*AGE	13.335	0.001271 **
WF*AGE	WF*WCM + WF*AGE	WF*WCM	19.363	<.001

WCM	WF + WCM + AGE	WF + AGE	1.2793	0.258
WF	WF + WCM + AGE	WCM + AGE	25.101	<.001 ***
AGE	WF + WCM + AGE	WF + WCM	15.025	0.0001061 ***

Note: The random effect structure for a given set of fixed effects will differ depending on whether it is considered as a main or comparison model

As can be seen from Table 4-11, of the three interactions, both the interaction between word frequency and word complexity and that between word frequency and age were found to explain significant unique variance. Of the three separate variables, both word frequency and age were found to explain significant unique variance.

We report the coefficients for a model containing an interaction of word frequency and word complexity, and an interaction of word frequency and age in Table 4-12. Random effects of participant, word position (adjective or noun) and item (phrase) on the intercept and a random effect of participant on all slopes (excluding age) were also included in the model.

Table 4-12 Coefficients for the Word Frequency*WCM (complexity) + Age model

Predictor	Parameter estimates			
	Log-odds	S.E	z value	p
Intercept	1.70520	0.13476	12.654	<.001 ***
Word Freq	0.21239	0.04879	4.353	<.001 ***
WCM	0.02487	0.04718	0.527	0.598148
Age	0.22697	0.05781	3.926	<.001 ***
Word Freq*WCM	-0.14893	0.04418	-3.371	0.000749 ***
Word Freq*Age	-0.07081	0.02969	-2.385	0.017090 *

We take the two analyses together as evidence for a main effect of word frequency, a main effect of age and for a word frequency by complexity interaction. However, the significant age by word frequency interaction indicate that these results are not completely robust to annotator disagreements. This needs to be confirmed in further studies.

4.11 Interim Discussion

The purpose of this experiment was to explore the effect of age, word frequency and word complexity (and their interactions) on production for the combined data. We predicted that there would be a main effect of age, where older children would be more accurate when producing phonemes in comparison to younger children. There was a significant positive main effect of age which supports this prediction. We also predicted that the negative effect of WCM on children's ability to repeat phonemes contained in each word will be reduced for older children. There was no significant complexity by age interaction. Our next prediction was that the positive effect of word frequency on children's ability to repeat phonemes contained in each word will be reduced for younger children. There was no significant word frequency by age interaction, so this prediction is not supported. Finally, we predicted that the negative effect of WCM on children's ability to repeat phonemes contained in each word will be reduced when that word is high frequency or part of a high frequency phrase. We found the opposite to be true, as was the case for the error data in both 3-year-olds and 4-year-olds.

4.12 Duration data

To explore the effect of age, word frequency and complexity on production in children, we measured the duration of productions for phrases which were produced correctly, using the recordings from both the 3-year-old and 4-year-old error data sets.

We aim to test the following hypotheses:

1. Older children will take less time to produce phrases in comparison to younger children
2. The effect of word complexity on children's speed of production will be reduced in older children compared to younger children
3. The effect of word frequency on children's speed of production will be reduced in younger children compared to older children
4. The effect of word complexity on children's speed of production will be reduced when that word is high frequency or part of a high frequency phrase

Our goal here was to test hypothesis 1 by seeing whether there is the predicted effect of age (a significant negative slope would indicate that duration increased for younger children even once length in phonemes was taken into account), hypothesis 2 by seeing whether there is the predicted interaction between age and word complexity (a significant negative term would indicate that older children had shorter durations for more complex items even once length in phonemes was taken into account), hypothesis 3 by seeing whether there is the predicted interaction between word frequency and age (a significant negative term would indicate older children had shorter durations for more frequent words frequencies even once length in phonemes was taken into account), and hypothesis 4 by seeing whether there is the predicted interaction between word complexity and our frequency predictors (a significant negative term for the interaction indicating that the effect of word complexity is lower for higher frequency words even once length in phonemes was taken into account).

4.13 Results

To address our research questions, we built a series of multilevel linear regression models. We started by building a model including word frequency, word complexity, age and their relevant 2-way interactions as fixed effects. Phoneme count was also included as a covariate. Random effects of participant, phoneme count, part of phrase (adjective or noun) and item (phrase) on the intercept and a random effect of participant on fixed effects were also included in the model. We tested these hypotheses using likelihood ratio tests in an iterative drop-one analysis, much like in the previous experiments.

Table 4-13 Fixed effects used in model comparisons

Term Evaluated	Larger Model	Reduced model	ChiSq	P value
WF*WCM	WF*WCM + WF*AGE + WCM*AGE	WF*AGE + WCM*AGE	6.3704	0.0116 *
WF*AGE	WF*WCM + WF*AGE + WCM*AGE	WF*WCM + WCM*AGE	0.1588	0.6902
WCM*AGE	WF*WCM + WF*AGE + WCM*AGE	WF*WCM + WF*AGE	0.5153	0.4729
WF*WCM	WF*WCM + AGE	AGE	74.716	<.001 ***
AGE	WF*WCM + AGE	WF*WCM	7.2093	0.007253 **
WCM	WF + WCM + AGE	WF + AGE	0.0851	0.7704
WF	WF + WCM + AGE	WCM + AGE	55.192	<.001 ***
AGE	WF + WCM + AGE	WF + WCM	1.1711	0.2792

Note: The random effect structure for a given set of fixed effects will differ depending on whether it is considered as a main or comparison model

As can be seen from Table 4-13, of the three interactions, the interaction between word frequency and word complexity was found to explain significant unique variance. Of the three separate variables, only word frequency was found to explain significant unique variance.

These model comparisons suggest that the word frequency by complexity interaction explains unique variance along with the word frequency term. We finally compared the WF*WCM model to a model including only word frequency (therefore removing WCM) and found that the interaction model (WF*WCM) was a significantly better fit than the comparison model ($\chi2(2) = 6.39$, $p= 0.041$). Taken together, we take these results as evidence that there is an effect of word frequency, but one that varies depending on the levels of complexity.

We report below the coefficients for a model containing an interaction of word frequency and word complexity in Table 4-14. Random effects of participant, word position (adjective or noun) and item (phrase) on the intercept and a random effect of participant on all slopes (excluding age) were also included in the model.

Table 4-14 Coefficients for the Word Frequency*WCM model

Predictor	Parameter estimates			
	Estimate	S.E	t value	p
Intercept	0.503701	0.068212	7.384	0.070850
Word Freq	-0.058667	0.007237	-8.107	<.001 ***
WCM	0.008983	0.009965	0.901	0.368168
Phoneme count	0.032823	0.009101	3.607	0.000395 ***
Word Freq*WCM	-0.015418	0.006106	-2.525	0.011903 *

Table 4-14 shows that there is a significant main effect of word frequency. The negative estimate value indicates that there is a significant negative slope, where duration was shorter for words with higher frequencies. There is also a significant word frequency by complexity interaction. The negative estimate value indicates that there is a significant negative term for the interaction, where the effect of word complexity is lower for higher frequency words. So, duration was shorter for high complexity words which were also high in word frequency.

4.13.1 Second coding

To check that the results we report above are not contingent on coder decisions about which there was not agreement in the error data, we re-ran the relevant analyses with the combined 3- and 4-year-old alternative datasets created above (see section 3.3.1 and 4.3.1). Age in months was added as a continuous variable and analyses were run on the phrases which were produced correctly in the alternative dataset.

Table 4-15 Fixed effects used in model comparisons

Term Evaluated	Larger Model	Reduced model	ChiSq	P value
WF*WCM	WF*WCM + WF*AGE + WCM*AGE	WF*AGE + WCM*AGE	4.3043	0.03802 *
WF*AGE	WF*WCM + WF*AGE + WCM*AGE	WF*WCM + WCM*AGE	0.0014	0.9707
WCM*AGE	WF*WCM + WF*AGE + WCM*AGE	WF*WCM + WF*AGE	0.7174	0.397
WF*WCM	WF*WCM + AGE	AGE	62.847	<.001 ***
AGE	WF*WCM + AGE	WF*WCM	9.9523	0.001606 **
WCM	WF + WCM + AGE	WF + AGE	0.9277	0.3355
WF	WF + WCM + AGE	WCM + AGE	59.133	<.001 ***
AGE	WF + WCM + AGE	WF + WCM	9.6892	0.001854 **

As can be seen from Table 4-15, of the three interactions, the interaction between word frequency and word complexity was found to explain significant unique variance. Of the three separate variables, both word frequency and age were found to explain significant unique variance.

We report below the coefficients for a model containing an interaction of word frequency and word complexity, and a main effect of age in Table 4-16. Random effects of participant, word position (adjective or noun) and item (phrase) on the intercept and a random effect of participant on all slopes (excluding age) were also included in the model.

Table 4-16 Coefficients for the Word Frequency*WCM model

Predictor	Parameter estimates			
	Estimate	S.E	t value	p
Intercept	0.505934	0.062892	8.045	0.062021
Word Freq	-0.054934	0.006895	-7.967	<.001 ***
WCM	0.010680	0.008877	1.203	0.229856
Age	0.022413	0.006940	3.230	0.001647 **
Phoneme count	0.027962	0.007613	3.673	0.000306 ***
Word Freq*WCM	-0.011845	0.005654	-2.095	0.036661 *

We take the two analyses together as further evidence for a main effect of word frequency and for a word frequency by complexity interaction. However, the main effect of age indicates that these results are not completely robust to annotator disagreements. This needs to be confirmed in further studies.

4.14 Interim discussion

The purpose of this experiment was to explore the effect of age, word frequency, word complexity (and their interactions) on the duration of production for the combined data. We predicted that older children would take less time to produce phrases in comparison to younger children. However, we found no main effect of age.

We also predicted that the effect of word complexity on children's speed of production would be reduced in older children compared to younger children. However, there was no significant complexity by age interaction. We next predicted that older children would take less time to repeat high frequency words in comparison to younger children. However, there was no significant word frequency by age interaction.

Finally, to further support the findings of experiment 1 and 2, we predicted that the effect of word complexity on children's speed of production would be reduced when that word is high frequency or part of a high frequency phrase (as was shown for the 4-year-old data). We found a significant negative word frequency by complexity interaction. Therefore, this prediction was supported. As for the 4-year-old data, children took less time to produce phrases containing words which were high in frequency when they were also high in complexity. This finding provides evidence of phonological/motor chunking.

4.15 Discussion

The purpose of this experiment was to explore the effect of age on fluency. To do this, we built models which included main effects of age, word frequency, word complexity (and their interactions) on the accuracy of phoneme production, and duration of phrase production.

We found that, overall, children made fewer errors when producing phrases containing high frequency words, and that this effect was greater when the phrases were low in complexity. However, the combined data also showed that older children made fewer errors when producing phonemes in general. It is possible, as for the 3-year-olds, that the frequency effect here is the result of phonological/motor chunking, but that this only occurs for the most phonologically simple words at this age. However, as for the 3-year-old data we cannot rule out the possibility that it is the result of an earlier pre-formulation stage of production.

More compelling evidence of phonological/motor chunking came from the duration analysis. We found that, overall, children were better at producing complex phrases when they contained high frequency words in line with our original prediction (which we did not find support for in the accuracy analysis).

Finally, as for the 3-year-old and 4-year-old data, there was no effect of phrase frequency in our sentence repetition task for these 2-word combinations. To further explore these findings, in the next chapter we will repeat this experiment using the same stimuli with adults.

5 Chapter 5 - Assessing the effect of word frequency, phrase frequency and phonological complexity on fluency in adults

5.1 Introduction

Adults are more fluent in their production of words and phrases which are more frequent in the input. Speakers take less time to name more frequent words in picture naming tasks (e.g., Jescheniak & Levelt, 1994) and make fewer errors for frequent words during naturalistic speech (e.g., Harley & MacAndrew, 2001) and during elicited imitation tasks (e.g., Vitevich and Sommers, 2003). Furthermore, frequent phrases are named more quickly in picture naming tasks (Janssen & Barber, 2012) and elicited imitation tasks (Arnon & Cohen Priva, 2013). One possible explanation as to why adults are better at producing words which are more frequent in language is phonological/motor chunking (e.g., Guenther, 2016). By this account, the production system stores frequent sequences as chunks in working memory. This allows for reduced processing, which means that these sequences are retrieved more quickly. In addition to this, motor routines for these frequent sequences are practiced more often and therefore become more efficient.

In the previous chapters we explored whether frequency effects in 3-year-old and 4-year-old children are due to phonological/motor chunking. To do this we used an elicited imitation task where we crossed word frequency, phrase frequency and complexity over 24 adjective-noun combinations. If children do use phonological/motor chunking, we would expect them to find words that are more complex easier to say after practice. Therefore, we predicted that for complex words, children would make fewer errors when producing phrases that were high in frequency or contained high frequency words, compared to complex phrases which were low in frequency or contained low frequency words. We also predicted that children would make very few errors when producing low complexity words – regardless of frequency. However, we found that both 3-year-olds and 4-year-olds made fewer errors when producing phrases which contained words which were high in frequency, but only when they were also low in complexity, while low frequency words were unaffected by complexity.

These findings suggest that there is an effect of word frequency occurring during an earlier stage in production. It looks like children do have a more robust knowledge of high frequency words and find it more difficult to produce these words when they are also high in complexity, compared to the low complexity words. However, we also found that older children made fewer errors in general in comparison to the younger children. Therefore, it is possible that the mechanisms underlying production are more robust as children get older, given that less errors were made.

We also predicted that children would take less time to produce complex phrases that were high in frequency, or contained high frequency words, compared to complex phrases that were low in frequency or contained low frequency words. We found that 3-year-olds produced phrases containing high frequency words more quickly, regardless of complexity. Whereas 4-year-olds produced phrases containing high frequency words more quickly when they were high in complexity – which is in line with our prediction. By measuring the duration of phrases which were produced correctly, this result suggests that once we remove the retrieval issue, children are better at producing complex phrases that have been practiced more often. Therefore, once children have robust conceptual and phonological knowledge of the target phrase, we see the frequency driven effect of phonological/motor chunking.

In the research described in this chapter, we aimed to explore how adults would perform in the same experiment, given that we would expect adults to have full conceptual and phonological knowledge of the target phrases. We are particularly interested in whether adults will use phonological/motor chunking for entire phrases. Although 3 and 4-year-olds did not show an effect of phrase frequency in our sentence repetition task for these 2-word combinations, it is possible that this effect occurs later in development.

To do this, we tested adults using the same methodology as for the previous experiments. Speakers were asked to repeat the same 24 adjective-noun phrases manipulated for phrase frequency, word frequency and word complexity. However, since we assume adults have robust knowledge of phrases, we expect the speakers to produce all the target items correctly. Therefore, we chose not to measure the accuracy of phoneme production, since we expected errors to be too rare to give meaningful variance. Instead, we measured the duration of correct productions.

We aimed to test the same hypotheses as for the duration data in experiment 1 and experiment 2:

1) Adults will take less time to repeat low complexity words.
2) Adults will take less time to repeat high frequency words and phrases.
3) The effect of word complexity on adult's speed of production will be reduced when that word is high frequency or part of a high frequency phrase.

We also again test the secondary prediction that phrase and word frequency will affect (promote or suppress) each other's effect on accuracy.

5.2 Method

5.2.1 Participants

59 monolingual English-speaking adults were recruited and included in analyses. One additional participant was recruited and excluded because they failed to produce at least 18 codable responses (75% of the items presented).

5.2.2 Design

This study used a within-subjects design. Our dependent variable was the duration of production for the entire phrase. Our independent variables were the same as for the previous experiments.

5.2.3 Materials

The materials were the same as for experiment one (see section 3.2.3)

5.2.4 Procedure

The procedure was the same as for experiment one (see section 3.2.4).

5.3 Transcription and coding

The recordings were coded by the author. Recordings were opened in Praat (Boersma & Weenink, 2018) where length of production was measured. The productions were coded for word and the onset and end of each word was recorded. To calculate the duration of production for each phrase, the word lengths were combined. Phrases were only coded for duration if produced correctly. Incorrect productions and productions that overlapped with the target recording were excluded (this meant that 2.75% of total productions were excluded, see section 2.2.3 for further details on coding).

5.4 Results

To address our research questions, we built a series of multilevel linear regression models. We started by building a model including phrase frequency, word frequency, word complexity and their relevant 2-way interactions as fixed effects. Phoneme count was also included as a covariate. Random effects of participant, part of phrase (adjective or noun) and item (phrase) on the intercept and a random effect of participant on all fixed effects were also included in the model.

Our goal here was to test hypothesis 1 by seeing whether there is the predicted effect of word complexity (a significant positive slope indicating that duration was higher for more complex items), hypothesis 2 by seeing whether there is the predicted effect of word and phrase frequency (a significant negative slope indicates that duration was greater for lower frequency words and phrases), and hypothesis 3 by seeing whether there is the predicted interaction between word complexity and our frequency predictors (a significant negative term for the interaction indicating that the effect of word complexity is lower for higher frequency words). This analysis procedure also allowed us to test the direction-agnostic prediction that word and phrase frequency would affect each other's effect on accuracy (a significant term in either direction for the interaction between word frequency and phrase frequency).

We tested these hypotheses using likelihood ratio tests in an iterative drop-one analysis. We first look at the effect of any interactions by building a full model and removing each interaction separately. Having removed any non-significant interactions but keeping all individual variables, we then perform a drop-one analysis on the remaining terms. We then remove any non-significant variables and repeat until only variables that explain significant variance, along with any component terms where interactions are included. Finally, to look at the contribution of the individual variables, we remove all remaining interactions where appropriate and perform a drop-one analysis on a model containing only the individual model terms.

Table 5-1 Fixed effects used in model comparisons

Term Evaluated	Larger Model	Reduced model	ChiSq	P value
WF*PF	WF*PF + WF*WCM + PF*WCM	WF*WCM + PF*WCM	32.178	<.001 ***
WF*WCM	WF*PF + WF*WCM + PF*WCM	WF*PF + PF*WCM	74.265	<.001 ***
PF*WCM	WF*PF + WF*WCM + PF*WCM	WF*PF + WF*WCM	3.8313	0.0503
WF*PF	WF*PF + WF*WCM	PF + WF*WCM	28.346	<.001 ***
WF*WCM	WF*PF + WF*WCM	WF*PF + WCM	78.14	<.001 ***
WCM	WF + PF + WCM	WF + PF	84.335	<.001 ***
PF	WF + PF + WCM	WF + WCM	1.5036	0.2201
WF	WF + PF + WCM	PF + WCM	137.33	<.001 ***

Note: The random effect structure for a given set of fixed effects will differ depending on whether it is considered as a larger or reduced model

As can be seen from Table 5-1, of the three interactions, the interaction between word frequency and word complexity, and that between word frequency and phrase frequency were found to explain significant unique variance. Of the three separate variables, both word frequency and word complexity were found to explain significant unique variance.

Taken together, we take these results as evidence that there is an interaction between word frequency and complexity and that there is an effect of word frequency, but one that varies depending on the frequency of the phrase frequency. We finally compared the WF*PF model to a model including only word frequency (therefore removing phrase frequency) and found that the interaction model (WF*PF) was a significantly better fit than the comparison model ($\chi2(2) = 7.05$, $p= 0.03$).

We report the coefficients for a model containing an interaction of word frequency and word complexity, and an interaction of word frequency and phrase frequency in Table 5-2. Random effects of participant, word position (adjective or noun) and item (phrase) on the intercept and a random effect of participant on all slopes were also included in the model.

Table 5-2 Coefficients for the full Word Frequency*WCM and Word Frequency*Phrase Frequency interaction model

Predictor	Parameter estimates			
	Estimate	S.E	t value	p
Intercept	0.4713	0.05066	9.305	0.0595
Word Freq	-0.03972	0.003668	-10.829	<.001 ***
Phrase Freq	0.02004	0.009098	2.203	0.0384 *
WCM	0.04752	0.004645	10.206	<.001 ***
Phoneme count	0.00006634	0.003801	0.017	0.9861
Word Freq*Phrase Freq	-0.01858	0.003451	-5.383	<.001 ***
Word Freq*WCM	-0.02771	0.003102	-8.932	<.001 ***

Table 5-2 shows that there is a significant main effect of word frequency. The negative estimate value indicates that there is a significant negative slope, where duration was shorter for words with higher frequencies. There is a significant main effect of phrase frequency. The positive estimate value indicates that there is a significant positive slope, where duration was longer for phrases with higher frequencies. There is also a significant main effect of complexity, where the positive estimate value indicates that there is a significant positive slope, where duration was longer for more complex words. There is a significant interaction between word frequency and phrase frequency. The negative estimate value indicates that the effect of phrase frequency is reduced for more frequent words. This suggests that when words are low in frequency, high frequency phrases take longer to produce. However, when words are high in frequency, phrase frequency does not impact speed of production. Finally, there is a significant interaction between word frequency and complexity. The negative estimate value indicates that there is a significant negative term for the interaction, where the effect of word complexity is lower for higher frequency words. So, duration was shorter for high complexity words which were also high in word frequency.

5.5 Discussion

The purpose of this experiment was to explore the effect of phrase frequency, word frequency, word complexity (and their interactions) on the duration of phrase production in adults. We are interested in how frequency effects compare to those in the child data. To explore this, we built models which included main effects of phrase frequency, word frequency, complexity, and their relevant interactions.

We predicted that adults would take less time to repeat low complexity words. We found a significant positive main effect of complexity which supports this prediction. We next predicted that adults would take less time to repeat high frequency words. We found a significant negative main effect of word frequency which supports this prediction. We also predicted that the effect of complexity on adult’s speed of production would be reduced when that word is high frequency or part of a high frequency phrase. An interaction between word/phrase frequency and complexity in this direction would provide evidence of adults undergoing phonological/motor chunking, where phrases/words become easier to say after practice. We found a significant negative interaction between complexity and word frequency which supports this prediction. Therefore, adults take less time to produce complex phrases when they contain words high in frequency. This finding suggests that adults use phonological/motor chunking and replicates our finding in the older children.

We also predicted that adults would take less time to repeat high frequency phrases. We found a significant positive main effect of phrase frequency, which suggests that adults took more time to produce phrases high in frequency. Finally, we tested the secondary prediction that phrase and word frequency would affect (promote or suppress) each other's effect on accuracy. We found a significant negative interaction between word frequency and phrase frequency which suggests that when words are low in frequency, high frequency phrases take longer to produce than low frequency phrases. However, when words are high in frequency, phrase frequency does not impact speed of production.

This finding is interesting because it suggests that when phrase frequency is high, and word frequency is low, the production of words is impeded. Speakers find words that are low in word frequency much more difficult to produce when they are also high in phrase frequency, compared to words which are low in word frequency and low in phrase frequency. One possible explanation for this finding is that there is competition between the representation of the word and of the phrase. Sosa and MacFarlane (2002) found evidence for competition between the representations of words and the representations of phrases. The authors measured reaction times to the word 'of' contained in collocations (e.g., 'part of') manipulated for frequency and found that response latencies increased when the collocation was high in frequency.

However, our findings demonstrate the opposite effect, where the production of words is impeded when phrase frequency is high, and word frequency is low. This suggests that speakers struggle to access the representations of low frequency words even more so when they are contained in high frequency phrases, which results in a slower production of the phrase. These opposing findings might reflect the fact that we are considering adjective-noun phrases instead of collocations. Furthermore, Sosa and MacFarlane do not manipulate word frequency in their experiment and only consider collocations which contain the word 'of.' This is considered to be a high frequency word because it is a preposition. Therefore, the authors do not consider collocations containing a low frequency word. It is not clear whether our results would have been replicated in this instance. Finally, the authors' findings only reflect the effects for one word. Therefore, more work is needed to explore these competition effects.

Nevertheless, the interaction effect between word frequency and phrase frequency suggests that word frequency is very important to the fluent production of words, whereas phrase frequency is not. This is reflected in the seen production advantage for high frequency words which is not reflected for high frequency phrases. This idea is further reinforced by the significant interaction between complexity and word frequency which suggests that motor learning is driven at the word frequency level, and not at the phrase frequency level. However, to fully ascertain what is driving this effect, future research is required.

In the next chapter we aim to explore the effect of word frequency and phrase frequency on accuracy of production in adults using artificial stimuli. Since adults are being taught novel strings, this experiment can be used to compare results to the child data. In addition to this, using artificial stimuli to measure error rate will rule out any effect of frequency being due to semantic related differences. It is possible that the previously seen word frequency effect, which we take as evidence that children have a more robust conceptual knowledge of high frequency words, is driven by systematic frequency-related differences in the meaning of words and phrases - so that speakers might have more robust conceptual knowledge of words which have stronger meanings, instead of this effect being driven by production frequency.

6 Chapter 6 - Assessing the effect of word frequency and phrase frequency using non-words in adults

6.1 Introduction

In the previous chapters we measured the effect of phrase frequency, word frequency, complexity, and their relevant 2-way interactions on the production of real phrases in 3-year-olds, 4-year-olds, and adults. The focus of these experiments was to explore whether children, like adults, use phonological/motor chunking.

We found that 3- and 4-year-old children were better at producing phrases when they contained words that were low in complexity, but only when the words were also high in frequency. In addition to this, low frequency words were seemingly unaffected by complexity. This finding suggests that children have a more robust knowledge of high frequency words and are better able to retrieve a lemma, for such words, but then find it more difficult to produce the high complexity words, in comparison to the low complexity words. From these results, it could be argued that phonological/motor chunking is not taking place in children at this age. If chunking had been taking place, we would expect phrases/words to become easier to say after practice. There would be a production advantage for complex phrases when they were part of a high frequency phrase or contained high frequency words, indicating that motor routines have become more efficient after practice.

We interpret these findings as children having more robust knowledge of high frequency words during the conceptualisation stage of production. However, it is also possible that this effect could be due to systematic frequency-related differences in the meaning of words and phrases. Speakers might have more robust conceptual knowledge of words which have stronger meanings, instead of this effect being driven by production frequency. This idea is most often explained in terms of a spreading activation network (e.g., Dell et al., 1997) where semantically related words are activated during the conceptualisation stage of production. This is supported by evidence of semantically related words being primed (e.g., Dell & O'Seaghdha, 1992).

One way to measure the strength of meanings is to look at imageability. Harley and MacAndrew (1992) looked at word substitution errors in naturalistic speech and found that high-imageability items have higher activation levels than low-imageability items (low-imageability items were more likely to be replaced with high-imageability items). Harley and MacAndrew (2001) later explored possible predictors of phonological errors which occur in naturalistic speech and found that high imageability words with a large number of close semantic neighbours have less accessible phonological forms and are often replaced by words which are more imageable. The authors take this as evidence of more imageable words having more highly activated competing items. However, this effect only stands when the words are also low in frequency.

However, in the previous experiments we also measured the duration for productions of phrases that were produced correctly and found that overall, children produced complex words that were also high frequency more quickly. These results are in line with our original prediction (which we did not find support for in the accuracy analysis). This suggests that once we remove the retrieval issue, looking only at phrases where children have robust conceptual and phonological knowledge, children are better at producing phrases which are high in word frequency and complexity. Therefore, frequency is driving phonological/motor chunking for these items.

We next measured the production duration in adults using the same stimuli and found a significant interaction between complexity and word frequency, where speakers took less time to produce complex words when they were also high in frequency, providing further evidence for adults using phonological/motor chunking. An additional and unexpected effect of phrase frequency was also observed, where phrases containing low frequency words took more time to produce when they were also high in phrase frequency, whereas when words were high in frequency, phrase frequency did not impact the speed of production. One interpretation of these findings is that word frequency is very important to the fluent production of words, while phrase frequency is not. This is reflected in the production advantage for high frequency words which is not seen for high frequency phrases. This suggests that motor learning is driven at the word frequency level, and not at the phrase frequency level.

It is possible that the representations of target words and phrases compete during production. Sosa and MacFarlane (2002) used a word monitoring task to measure adult reaction times to the word ‘of’ for collocations (e.g., ‘part of’) which varied in frequency. It was found that adults took longer to respond to ‘of’ when the collocation was higher in frequency. The authors argued that holistic storage of the collocation hindered access to the word ‘of’. In our case, it might be that speakers struggle to access the representations of low frequency words even more so when they are contained in high frequency phrases. This conflict is not seen when the words are themselves higher frequency.

The current experiment explored the effect of word frequency and phrase frequency (and their interaction) on the accuracy of phoneme production in adults using non-word phrases. We aim to explore whether the previously seen word frequency effect is driven by systematic frequency-related differences in the meaning of words and phrases, where speakers might have more robust conceptual knowledge of words which have stronger meanings, instead of this effect being driven by production frequency. To do this we use artificial stimuli to rule out any effect of frequency being due to semantic related differences. Furthermore, since adults are being taught novel strings, this experiment can possibly be used to compare results to the child data. However, adult speakers do not have any knowledge of these non-words before training, while children’s knowledge of real words is likely to be partial (stronger or weaker on a continuum). Therefore, any comparison between results should consider this.

We are interested in whether word frequency effects will drive fluency, as seen in both children and adults in the previous experiments. We are also interested in whether there will be an interaction between word and phrase frequency for newly learned words.

Participants were asked to repeat non-word phrases which were fully crossed for phrase frequency and word frequency. 8 non-words containing 3 syllables (9 phonemes) with the same CCVCCVC structure were used. Each phrase consisted of 2 non-words (e.g., "*kloobroapait twoadroosig*") so that we had 4 non-word phrases. All non-words were legal in English but had a low probability of occurring in language. The non-word phrases were manipulated for phrase frequency and word frequency and were presented via a training stage. During this training stage high frequency phrases were presented to the speaker more often than phrases considered to be low frequency. Words considered to be high frequency were presented more times than words considered to be low frequency and the number of times words were seen within a phrase were balanced with their manipulated word frequency band (see section 6.4). Once this training stage was completed, the test phase immediately began. For these experimental trials, the speaker produced each target phrase once after being presented with the auditory model. The procedure continued until all 4 test phrases were produced. Finally, their productions of the 4 test phrases were coded phoneme by phoneme (as in the previous experiments).

We test the following hypotheses:

1. Participant accuracy in producing target phonemes will be higher when those phonemes are part of phrases that they repeated frequently in a preceding training phase than when they are part of phrases they repeated infrequently.

2. Participant accuracy in producing target phonemes will be higher when those phonemes are part of words that they repeated frequently in a preceding training phase than when they are part of words they repeated infrequently.

3. We also test the secondary prediction that phrase and word frequency will affect (promote or suppress) each other's effect on accuracy.

6.2 Method

This study was preregistered on Open Science Framework. URL: https://osf.io/sur7h

6.3 Experiment 4

6.3.1 Participants

48 adult speakers with English as their first language were recruited and included in analysis for each study. Each participant was paid a £5 Love2Shop voucher for taking part. All speakers were tested in a quiet lab in the University of Liverpool. 7 participants who copied fewer than 95% of trials, did not attempt to produce a test phrase or repeated the test item before the item had been produced were excluded and a new participant was recruited to replace them.

A sample size of 48 was determined in advance (and preregistered) based upon blocks of 24 participants needed for counterbalancing (see Table 6-3). This will allow each of the 24 unique sets of stimuli to be presented to two participants each. A power analysis established that 48 participants was above the minimum required. We conducted a simulation-based power analysis using SIMR (Green & MacLeod, 2016). The simulation used a mixed effects logistic regression model with a participant-varying intercept and item-varying intercept and slope. It assumed a small effect size (log-odds of 0.363 for effect of condition; equivalent to Cohen's d=0.2). For the purposes of simulation, an intercept of 0.405 was assumed (based on a correct response rate of 60% in the low frequency condition), and random effect terms equivalent to a standard deviations of 1/3 of the varying parameter (intercept or slope) in all cases. A power of .8. (in detecting a simple effect of frequency) was observed to be achievable with alpha=0.05 with a sample of 37 participants.

6.3.2 Design

This study used a within-subjects design. Our dependent variable was a score of 1 (correctly produced) or 0 (not correctly produced) for each phoneme in the target items. See section 2.2.2.2 for more details.

Our independent variables were as follows:

- Phrase Frequency – the manipulated frequency of target non-word phrases (the number of times the participant is required to repeat a target non-word phrase during the training phase of the experiment, see Table 6-3).

- Word Frequency – the manipulated frequency of target non-words (the number of times the participant is required to repeat a target non-word during the training phase of the experiment, see Table 6-3).

6.3.3 Materials

8 non-words containing 3 syllables (9 phonemes), and all with the same CCVCCVC structure were constructed. To make coding straightforward, each non-word was given a stop-consonant for the first and final phoneme (designed for measuring production onset and duration). Consonant clusters were included to increase complexity while biphone probability was held constant (see Table 6-1). All non-words are legal in English - all bi-phone probabilities are more than 0 but have a low probability of occurring. Bi-phone probabilities were calculated using the Phonotactic Probability Calculator (PPC: Vitevitch, & Luce, 2004), where bi-phone probability is calculated according to the probability of neighbouring phonemes occurring for that specific position in the non-word. The highest biphone probability for PCC was 0.0081 and the lowest biphone probability was 0.0001.

Since the PPC is an American calculator, the position independent transitional probabilities for each biphone were also calculated using a British calculator in order to ensure that all bi-phones had a score greater than than 0 (Moreland, 2011). Here the highest transitional probability was 0.2092 and the lowest transitional probability was 0.0027. Phoneme count was also held constant over all non-words. Target phrases were recorded by a female English speaker with an East England regional accent. Stress was placed on the first syllable. Volume was normalised and all phrases recordings were adjusted to be the same length.

Table 6-1 List of non-words with orthographic representation

	Word	Orthographic representation	Total Biphone Probability (PCC)	Total Transitional Probability
A	klubropaɪt	kloobroapait	0.0138	0.0568
B	twodrusɪg	twoadroosig	0.0082	0.0353
C	draɪpwosæb	draipwoasab	0.0129	0.0250
D	pl^tlubig	plutloobeeg	0.0135	0.0312
E	graʊspɛdæk	growspedak	0.0141	0.0569
F	gwɪklaʊtaɪb	gwiklowtaib	0.0106	0.0402
G	dwɛgr^kit	dwegrukeet	0.0082	0.0493
H	bliskædɛp	bleeskadep	0.0113	0.0523

Table 6-2 List of non-words with biphone transition probability, position independent and transitional probabilities

Word	Phoneme	Biphone transition probability using PCC	Position independent transitional probabilities (Moreland, 2011)
klubropaɪt	k		
	l	0.0067	0.0457
	u	0.0025	0.0212
	b	0.0003	0.0391
	r	0.0006	0.1076
	o	0.0018	0.0484
	p	0.0005	0.0360
	aɪ	0.0002	0.0187
	t	0.0012	0.1112
twodrusɪg	t		
	w	0.0013	0.0087
	o	0.0014	0.0113
	d	0.0004	0.0423
	r	0.0019	0.0476
	u	0.0014	0.0301
	s	0.0004	0.0639
	ɪ	0.001	0.0913
	g	0.0004	0.0272
draɪpwosæb	d		
	r	0.0048	0.0476
	aɪ	0.0056	0.0424
	p	0.0003	0.0316
	w	0.0001	0.0027
	o	0.0003	0.0113
	s	0.001	0.0524
	æ	0.0001	0.0196

	b	0.0007	0.0335
pl^tlubig	p		
	l	0.006	0.0682
	^	0.0033	0.0155
	t	0.0004	0.0428
	l	0.001	0.0090
	u	0.0022	0.0212
	b	0.0002	0.0391
	i	0.0003	0.0597
	g	0.0001	0.0089
graʊspɛdæk	g		
	r	0.0081	0.1983
	aʊ	0.0018	0.0096
	s	0.0001	0.1108
	p	0.0015	0.0472
	ɛ	0.0008	0.0555
	d	0.0013	0.0604
	æ	0.0002	0.0182
	k	0.0003	0.1119
gwɪklaʊtaɪb	g		
	w	0.0001	0.0218
	ɪ	0.0039	0.2092
	k	0.0035	0.1374
	l	0.0014	0.0457
	aʊ	0.0001	0.0052
	t	0.0004	0.1831
	aɪ	0.0007	0.0203
	b	0.0005	0.0246
dwɛgr^kit	d		
	w	0.0003	0.0100
	ɛ	0.0026	0.0936
	g	0.0004	0.0200
	r	0.0026	0.1983
	^	0.001	0.0286
	k	0.0006	0.0624
	i	0.0002	0.0262
	t	0.0005	0.0334
bliskædɛp	b		
	l	0.005	0.0548
	i	0.0026	0.0858
	s	0.0009	0.0292
	k	0.002	0.0593
	æ	0.0004	0.0525
	d	0.0001	0.0394
	ɛ	0.0001	0.0420
	p	0.0002	0.0404

6.3.4 Procedure

This experiment used an elicited imitation paradigm and consisted of a training and test phase. The experimenter sat with the participant in front of a laptop. The participant was told that they would hear a recorded voice produce some made up words and phrases. They would need to listen to the recording and then repeat each non-word or phrase in turn. No help was given to the speaker and no phrases were repeated. This procedure continued until all practice targets were produced. The relative frequency of the different phrases was varied as explained below. Once this training stage was completed, the test phase immediately began. For these experimental trials, the speaker produced each target phrase once after being presented with the auditory model. The procedure continued until all 4 test phrases were produced.

6.4 Counterbalancing

The assigning of the different phrases to the different phrase and frequency bands was varied, such that all phrases occurred with all combinations of word and phrase frequency. For example, for one participant, the phrase “kloobroapait twoadroosig” occurred 8 times, and the two component words “kloobroapait” and “twoadroosig” occured as isolated words 8 times each, making this the high phrase frequency and high word frequency item, while for another participant that phrase occurred 8 times, and each of the words zero times making it the high phrase frequency, low word frequency, and so on for all four of the frequency band combinations. There were thus 4! = 24 unique sets of stimuli.

Table 6-3 Example training trials

Phrase	Phrase Frequency		Word	Word Frequency	
	Frequency	Repetitions		Frequency	Repetitions
AB	HIGH	8	A	HIGH	8
			B	HIGH	8
CD	HIGH	8	C	LOW	0
			D	LOW	0
EF	LOW	2	E	HIGH	14
			F	HIGH	14
GH	LOW	2	G	LOW	6
			H	LOW	6

6.5 Transcription and coding

The recordings were coded by the author. Recordings were opened in Praat (Boersma & Weenink, 2018) where the speaker’s phoneme productions were transcribed using the CELEX transcription alphabet, which is a machine-readable adaptation of the International Phonetic Alphabet. Next, each of the target phonemes was given a score of 0 or 1 depending on whether a match was found for that phoneme in the attempted production when the two sequences (target phoneme sequence and produced phoneme sequence) were aligned using the Needleman-Wunsch Algorithm (see section 2.2).

20% of recordings were transcribed by a second coder (see section2.2.2). The percentage of agreement and the Kappa score were calculated. Percentages of agreement for interrater reliability were 81.7% and the Kappa scores was 0.639 representing substantial agreement (Cohen, 1960).

6.6 Results:

To address our research questions, we built a series of multilevel logistic regression models. We started by building a model including phrase frequency, word frequency and their 2-way interaction as fixed effects. Random effects of participant, part of phrase (adjective or noun) and item (phrase) on the intercept and a random effect of participant on fixed effects were also included in the model.

Our goal here was to test hypothesis 1 by seeing whether there is the predicted effect of phrase frequency (a significant positive slope indicating that accuracy was higher for more frequent items), hypothesis 2 by seeing whether there is the predicted effect of word frequency (a significant positive slope indicating that accuracy was greater for more frequent words), and hypothesis 3 by seeing whether there is the predicted interaction between phrase and word frequency predictors (a significant term for the interaction in either direction).

We tested these hypotheses using likelihood ratio tests in an iterative drop-one analysis. We first look at the effect of the interaction by building a full model and removing each term separately. Next, to look at the contribution of the individual variables, we perform a drop-one analysis on a model containing only the individual model terms.

Table 6-4 Fixed effects used in model comparisons for all possible combinations

Term Evaluated	Larger Model	Reduced model	Chi Square	P value
Interaction	Word Freq*Phrase Freq	Word Freq + Phrase Freq	1.6563	0.1981
Phrase Freq	Word Freq*Phrase Freq	Word Freq	1.7017	0.427
Word Freq	Word Freq*Phrase Freq	Phrase Freq	7.8539	0.0197 *
Phrase Freq	Word Freq + Phrase Freq	Word Freq	0.0084	0.9269
Word Freq	Word Freq + Phrase Freq	Phrase Freq	7.4284	0.006385 **

Note: The random effect structure for a given set of fixed effects will differ depending on whether it is considered as a main or comparison model

As can be seen from Table 6-4, the interaction between word frequency and phrase frequency was not found to explain significant unique variance. Of the two separate variables, only word frequency was found to explain significant unique variance. Taken together we take these results as evidence that there is an effect of word frequency, but not of phrase frequency.

We report below the coefficients for a model containing only word frequency (see Table 6-5). Random effects of participant, word position (adjective or noun) and item (phrase) on the intercept and a random effect of participant on the interaction were also included in the model. The positive coefficient indicates that phrases were produced more accurately when they contained high frequency words.

Table 6-5 Coefficients for the Word Frequency model

Predictor	Parameter estimates		Wald's test	
	Logs-odds	S.E	Z	p
Intercept	0.37533	0.19911	1.885	0.05943
Word Frequency (HIGH)	0.22470	0.07958	2.824	0.00475 **

6.7 Discussion for Experiment 4

The purpose of this experiment was to explore the effect of phrase frequency, word frequency, and their interactions on the accuracy of phoneme production for artificial stimuli. We predicted that participant accuracy would be higher for phonemes part of phrases that they repeated frequently in a preceding training phase than when they are part of phrases they repeated infrequently. However, there was no main effect of phrase frequency.

We also predicted that participant accuracy would be higher for phonemes which were part of words that they repeated frequently in a preceding training phase than when they were part of words they repeated infrequently. We observed an effect of word frequency where speakers were more accurate in their production of words which were repeated more frequently in the training phase.

Finally, we predicted that there would be an interaction between phrase and word frequency effects, such that the increase in production accuracy seen for high frequency words relative to low frequency words will be reduced when the phrase is high frequency and vice versa. However, there was no interaction between word frequency and phrase frequency. These results suggest that speakers are engaging in motor learning at the word level but not at the phrase level.

As in the previous experiments, it is not clear whether this effect is being driven by phonological working memory or by more efficient practice of the articulatory system. Therefore, in the next experiment we aimed to explore the effect of phrase frequency and word frequency (and their 2-way interaction) on the production of phrases for which the speaker does not have to hold the phrase in memory.

6.8 Experiment 5

In the previous experiment we aimed to explore the effect of phrase frequency and word frequency on fluency in adults using artificial stimuli. We are particularly interested in whether adults engage in phonological/motor chunking for non-words. Since adults were taught novel strings, the results from the previous experiment could be compared to the results from the child data. We also aimed to explore whether the previously found effect of word frequency is instead due to systematic frequency-related differences in the meaning of words and phrases. We found a main effect of word frequency which suggests that previous findings were not due to semantic related differences. Instead, it supports the idea that phonological/motor chunking is driven by word frequency.

To complete the task in experiment 4, the speaker was required to hold the phrase in working memory when it was heard, before producing each word in turn. However, sentence repetition tasks which use real language assume that the speaker uses representations of words supported by long-term memory, where these representations are not purely phonological. This experiment required production of novel phrases which meant that the speaker could not rely on these representations supported by long-term memory. Therefore, we used the same task as in experiment 4, but this time included a visual aid to make the task more analogous to the real language tasks seen in the previous chapters. Here the non-words were presented on the screen which acts as a replacement for the participant's linguistic representations which bolster memory for the real words. Since the stimulus is displayed, the target phrase does not need to be held in memory.

Given that the speakers were not required to hold the stimuli in memory, we expected this task to be easier than experiment 4. To account for this, we halved the number of training trials and measured speaker's productions using the same methodology as in Experiment 1. Productions were transcribed by hand using CELEX. For each target phoneme a score of 0 or 1 was assigned depending on whether a match was found for that phoneme in the speaker's attempted production.

Our hypotheses were the same as for the previous experiment:

1. Participant accuracy in producing target phonemes will be higher when those phonemes are part of phrases that they repeated frequently in a preceding training phase than when they are part of phrases they repeated infrequently.

2. Participant accuracy in producing target phonemes will be higher when those phonemes are part of words that they repeated frequently in a preceding training phase than when they are part of words they repeated infrequently.

3. We also test the secondary prediction that phrase and word frequency will affect (promote or suppress) each other's effect on accuracy.

6.9 Method

This study was preregistered on Open Science Framework under the same preregistration as experiment 1. URL: https://osf.io/sur7h

6.9.1 Participants

48 adult speakers with English as their first language were recruited and included in analysis for each study. Each participant received a £5 Love2Shop voucher. All speakers were tested in a quiet lab in the University of Liverpool. 1 participant did not attempt to produce a test phrase and a new participant was recruited to replace them. 2 additional participants were excluded and replaced due to an experimental error where a test phrase was unintentionally skipped by the experimenter during the experiment.

A sample size of 48 was determined in advance (and preregistered) based upon blocks of 24 participants needed for counterbalancing. A power analysis established that 48 participants was above the minimum required (this was the same power analysis as in experiment 4, see section 6.3.1)

6.9.2 Materials

The same materials were used as for the previous study, but for this experiment phrases were presented with their orthographic representation on the screen (see table 6-1), while the auditory stimulus was played.

6.9.3 Procedure

The procedure was the same as in Experiment 1.

6.10 Transcription and coding

20% of recordings were transcribed by a second coder (see section 2.2.2). The percentage of agreement and the Kappa score were calculated. Percentages of agreement for interrater reliability were 84.3% and Kappa scores were 0.618 representing substantial agreement (Cohen, 1960).

6.11 Results

To address our research questions, we built a series of multilevel logistic regression models. We started by building a model including phrase frequency, word frequency and their relevant 2-way interactions as fixed effects. Random effects of participant, part of phrase (adjective or noun) and item (phrase) on the intercept and a random effect of participant on fixed effects were also included in the model.

Our goal here was to test hypothesis 1 by seeing whether there is the predicted effect of phrase frequency (a significant positive slope indicating that accuracy was higher for more frequent items), hypothesis 2 by seeing whether there is the predicted effect of word frequency (a significant positive slope indicating that accuracy was greater for more frequent words), and hypothesis 3 by seeing whether there is the predicted interaction between phrase and word frequency predictors (a significant term for the interaction in either direction). Like the previous experiment, we tested these hypotheses using likelihood ratio tests in an iterative drop-one analysis.

Table 6-6 Fixed effects used in model comparisons for all possible combinations.

Term Evaluated	Larger Model	Reduced model	ChiSquare	P value
Interaction	Word Freq*Phrase Freq	Word Freq + Phrase Freq	0.0464	0.8294
Phrase Freq	Word Freq*Phrase Freq	Word Freq	2.2579	0.3234
Word Freq	Word Freq*Phrase Freq	Phrase Freq	8.044	0.01792 *
Phrase Freq	Word Freq + Phrase Freq	Word Freq	1.9245	0.1654
Word Freq	Word Freq + Phrase Freq	Phrase Freq	8.0158	0.004637 **

Note: The random effect structure for a given set of fixed effects will differ depending on whether it is considered as a main or comparison model

As can be seen from Table 6-6, the interaction between word frequency and phrase frequency was not found to explain significant unique variance. Of the two separate variables, only word frequency was found to explain significant unique variance. Taken together we take these results as evidence that there is an effect of word frequency.

We report below the coefficients for a model containing only word frequency (see table 6-7). Random effects of participant, word position (adjective or noun) and item (phrase) on the intercept and a random effect of participant on the interaction were also included in the model. The positive coefficient indicates that phrases were produced more accurately when they contained high frequency words.

Table 6-7 Coefficients for the Word Frequency model

Predictor	Parameter estimates		Wald's test	
	Logs-odds	S.E	Z	p
Intercept	1.24924	0.16280	7.674	<.001***
Word Frequency (HIGH)	0.29413	0.09996	2.943	0.00325**

6.12 Discussion for Experiment 5

The purpose of this experiment was to explore the effect of phrase frequency, word frequency, and their interaction, on the accuracy of phoneme production for artificial stimuli. The speaker was presented with both the auditory and the orthographic representation of the non-word phrase, so that we could explore frequency effects for a more natural language task. We predicted that participant accuracy would be higher for phonemes part of phrases that they repeated frequently in a preceding training phase than when they are part of phrases they repeated infrequently. However, there was no main effect of phrase frequency.

We also predicted that participant accuracy in producing target phonemes would be higher when those phonemes were part of words that they repeated frequently in a preceding training phase, than when they are part of words they repeated infrequently. We found a significant positive main effect of word frequency. Therefore, this prediction is supported. Finally, we predicted that there would be an interaction between phrase and word frequency effects, such that the increase in production accuracy seen for high frequency words relative to low frequency words will be reduced when the phrase is high frequency and vice versa.

However, there was no significant interaction between word and phrase frequency. This prediction was not supported.

6.13 Discussion

The purpose of the research described in this chapter was to explore the effect of phrase frequency, word frequency, and their interactions, on the accuracy of phoneme production for artificial stimuli. We were particularly interested in whether word frequency effects would drive fluency, as seen in both children and adults in the previous experiments. We also aimed to explore whether the previously seen frequency effects were driven by a more robust knowledge of high frequency words during the conceptualisation stage of production, or whether this difference was due to systematic frequency-related differences in the meaning of words and phrases.

To do this, we conducted 2 separate experiments. In experiment 4, the speaker was presented with the auditory representation of the target phrases and was asked to repeat the phrase. In experiment 5, the speaker was presented with both the auditory and the orthographic representation of the non-word phrase and was asked to repeat the phrase. This meant that we explored the effect of frequency on error rate using artificial stimuli in experiment 4, before developing this task to be more analogous to real language tasks in experiment 5.

To complete the task in experiment 4, the speaker was required to hold the phrase in working memory when it was heard, before producing each word in turn. However, sentence repetition tasks which use real language assume that the speaker uses representations of words supported by long-term memory. In experiment 4, the speaker cannot rely on these representations supported by long-term memory because the phrases are novel. This means that the production process in this experiment was not natural and arguably cannot be compared to real life speech. Therefore, we used the same task as in experiment 4, but this time included a visual aid to make the task more analogous to the real language tasks seen in the previous chapters. Here the non-words were presented on the screen which acts as a replacement for the participant's linguistic representations which bolster memory for the real words. Since the stimulus is displayed, the target phrase does not need to be held in memory.

We found an effect of word frequency in both experiments, where speakers made fewer errors when producing phrases consisting of high frequency words. These findings suggest that word frequency is driving phonological/motor chunking during the formulation stage of lexical access. Guenther (2016) suggests that chunking at the formulation stage is driven by both phonological working memory (where motor programs are stored) and articulation (where practice of gestures leads to faster production).

Finally, there was no effect of phrase frequency in this experiment, as was the case for the child data. Given that this experiment used artificial stimuli, adults are being taught novel strings and so the results of this experiment can be directly compared to the child data. One interpretation of this finding is that, during development, word frequency is the main factor driving fluency effects. However, once speakers have robust conceptual and phonological knowledge, phrase frequency effects are possibly driven by long-term memory and redintegration.

7 Chapter 7 – General Discussion

The objective of this thesis was to explore the underlying processes that affect fluency in children. We explored whether the previous findings for independent effects of phrase frequency, word frequency and word complexity on phrase production held when using phrases that are simultaneously manipulated for all three factors. We also aimed to understand how the effects of these properties interact. To explore this, we asked both children and adults to repeat a series of phrases manipulated for phrase frequency, word frequency and phonological complexity, where error rate and duration of production were used as measures of fluency. In this chapter, each experiment is described in turn while the aims, methods used, results and implications are considered. We then present an informal account of production which considers the observed effects of word frequency, phrase frequency and complexity (and their interactions) on production for 2-word phrases, while also reflecting previous literature.

Both children (e.g., Sosa and Stoel-Gammon, 2012) and adults (e.g., Jescheniak & Levelt, 1994) show a processing advantage for the production of words which are high in frequency. Different models of production which aim to explain this effect generally accept that there is a semantically/syntactically specified representation followed by a phonologically specified representation. In some accounts this system is said to be organised by frequency, with more frequent words/sounds being more easily accessed. By these accounts, word frequency effects in adults are most commonly attributed to the second (formulation) stage of lexical access, where the phonological form of the word is realised (see Levelt et al, 1999). However, it is also possible that frequency effects might plausibly be attributed to another stage of production.

Alongside the word frequency effects discussed above, the last decade has seen a number of papers published showing that speakers are affected not just by the frequency of words, but also of phrases, such that children (e.g., Bannard & Matthews, 2008) and adults (e.g., Janssen & Barber, 2012) are better at producing high frequency phrases. The number of reported phrase frequency effects have led to the idea that frequent use of phrases leads to chunking, where chunked multi-word sequences have a processing advantage. However, it is less clear at which stage of lexical access phrase frequency effects might occur.

We aimed to explore the cognitive representations and mechanisms underlying both word and phrase frequency effects seen during production in children, and to further determine whether these differ from adults. If errors do come from the formulation stage, frequency-related effects in adult production are usually conceptualised, even if only implicitly, as an effect of phonological/motor sequence learning, with words that are difficult to articulate when first encountered becoming easier to say due to practice. Although models of production which consider chunking do not necessarily agree on the structure or format of chunks, they generally accept that there is some level of chunking which occurs during the phonological stage of production. Here, sequences are produced more fluently because they have been practiced more often in comparison to less frequent sequences.

Segawa et al. (2015) explored the tradeoff between phonological complexity and frequency by asking American English speakers to produce mono-syllabic non-words containing phonotactically legal and illegal consonant clusters. It was found that speakers had a significantly lower error rate after practice for the illegal sequences but did not differ in their repetition of legal syllables. Furthermore, after introducing novel illegal sequences, speakers made more errors during novel illegal syllables compared to learned illegal syllables. This latter finding provides evidence for learned motor programs - given that performance improvement was specific to the stimuli encountered, improvement could not be a result of the learning of phonological rules.

Therefore, we were interested in whether, like for adults, frequency effects in children are attributable to phonological/motor chunking. We explored whether main effects of word frequency, phrase frequency and complexity would hold when using phrases that are simultaneously manipulated for all three factors. Furthermore, if motor learning does occur, words should become easier to say after practice. Therefore, we would expect to see an interaction between phrase/word frequency and complexity, where more complex words/phrases are produced more fluently when they are also more frequent compared to complex words/phrases which are infrequent. In addition to this, our experiments are the first to cross factors of word and phrase frequency. We aimed to explore whether these effects are independent and, if not, whether they are complementary or if the accessing of representations at different levels might lead to competition.

7.1 Experiment 1

In the first experiment, children aged 3;2 to 3;9 repeated phrases manipulated for phrase frequency, word frequency and phonological complexity. All productions were coded phoneme by phoneme to provide us with a graded measure of accuracy for the repeated phrases. We additionally measured the duration of productions for phrases which were produced correctly.

Both children and adults show a processing advantage for the production of phrases and also words which are high in frequency, while speakers also find complex words more difficult to produce than words that are less complex. Therefore, we aimed to explore whether these main effects would hold when using phrases that are simultaneously manipulated for all three factors. However, it is also known that speakers do improve in their production of complex words when they are also more frequent (e.g., Segawa et al., 2015). One possible explanation for this is some kind of phonological/motor chunking. We were interested in whether, like for adults, frequency effects in children are attributable to this chunking. If motor learning does occur in 3-year-olds, words should become easier to say after practice. Furthermore, this is the first study to cross factors of word and phrase frequency. We aimed to explore whether these effects are independent and, if not, whether they are complementary or if the accessing of representations at different levels might lead to competition.

We hypothesised that there would be a main effect of word and phrase frequency, where speakers would be more fluent (reduced error rate and reduced duration) in their production of high frequency phrases and high frequency words, in comparison to low frequency phrases and low frequency words. We also predicted that there would be a main effect of word complexity, where children would make fewer errors and take less time to produce words which were less complex regardless of frequency. Finally, we predicted that the negative effect of word complexity on children's error rate and speed of production would be reduced when that word is high frequency or part of a high frequency phrase – so, children would be better at producing high complexity phrases when they were high in frequency or when they contained high frequency words (due to phonological/motor chunking).

We conducted multilevel regression analyses to explore our research questions. Firstly, all productions were coded phoneme by phoneme to provide us with a graded measure of accuracy for the repeated phrases. To perform this coding, recordings were opened in Praat (Boersma & Weenink, 2018) where the children's phoneme productions were transcribed. To conduct the statistical analyses for these error data, we first removed insertions from the dataset. An accuracy score was calculated phoneme by phoneme where matches were given a score of 1 and all other operations were given a score of 0. The accuracy score (0 and 1s) was used directly in a logistic regression. We found a main effect of word frequency where children made fewer errors when producing phrases which contained high frequency words, but there was no main effect of phrase frequency or complexity. In addition to this, we found that children made fewer errors when producing phrases which were low in complexity, but only when the phrases contained words which were high in frequency. This was the opposite of what was predicted.

For the duration analysis, the length of the coded phonemes was extracted from the phoneme-by phoneme coding. The individual phoneme durations were taken and combined to produce word durations. Here, we found a main effect of word frequency where children took less time to produce phrases containing high frequency words. We did not find a main effect of phrase frequency or complexity and no significant interaction was seen.

Taken together, these results suggest that children have more robust knowledge, or a more robust retrieval process, for high frequency words. Our results show that children were more accurate when producing phrases when they contained high frequency words that were low in complexity. It is possible that children might have more robust knowledge of, or a more robust retrieval process for, high frequency words. This indicates that there is an effect of frequency during lemma retrieval. Once these high frequency words have been retrieved, speakers find it more difficult to produce the high complexity words in comparison to the low complexity words. Furthermore, our results show that children did not differ in their error rate for words of varying complexity when they were low in frequency. This might be because children do not (at this age) have strong representations for the lemmas of words which are not frequent. Articulatory complexity it seems makes little difference when the speaker does not have knowledge of the lemmas that are required to begin the formulation process.

We next measured the duration rate for phrases produced correctly. This meant that we were measuring fluency in children who had robust conceptual and phonological knowledge for these phrases. Since a possible explanation for the above findings is that children have incomplete knowledge of low frequency forms, by measuring the duration of only the correct productions, we set aside this possibility of incomplete knowledge. Children must have full conceptual knowledge of the word, and robust enough phonological representations to produce these phrases correctly. This means that we can explore differences in the speed of production for phrases, so that any effect of frequency, and or complexity, seems more likely to be due to the second, phonological realisation stage of speech production.

We found a main effect of word frequency which suggests that speakers start to show an advantage for producing high frequency words, regardless of complexity. This pattern of results is not inconsistent with the claim that phonological/motor chunking is occurring. This is because we would expect speakers who do engage in chunking to find words easier to say after practice. This effect of word frequency shows a production advantage for words which are more frequent which is in line with this prediction. In addition to this, we take our finding as evidence of word frequency having an effect at the phonological realisation stage, which is the stage at which chunking is said to take place. However, this is not as strong evidence as if we had seen an interaction between frequency and complexity. A production advantage for complex phrases which were more frequent would show that the speaker has overcome the inherent complexity of the word due to practice, driven by chunking. No such pattern was seen here.

7.2 Experiment 2

In experiment 2, we replicated experiment 1 but in slightly older children. We did this to explore whether there was a developmental effect. It is possible that older children might have more robust conceptual and phonological knowledge in general, and so might have shown evidence of phonological/motor chunking which was not seen in the younger children. Children aged 3;9 to 4;4 repeated the same phrases as in experiment 1. As in experiment 1, our measures of fluency were error rate and the duration of production for correctly produced phrases. The hypotheses were the same as for experiment 1.

We conducted multilevel regression analyses to explore our research questions. For error rate, we observed the same results as for experiment 1. Children made fewer errors when producing phrases which contained high frequency words, but there was no main effect of phrase frequency or complexity. Again, we found that children made fewer errors when producing phrases which were low in complexity, but only when the phrases contained words which were high in frequency. When producing words which were low in frequency, they were seemingly unaffected by complexity.

In the duration analysis, we found that children took less time to produce phrases containing high frequency words. We did not find a main effect of phrase frequency or complexity. However, unlike in experiment 1, we found a word frequency by complexity interaction, so that children did take less time to produce high complexity phrases when they contained high frequency words. Once we remove the retrieval issue, for phrases where children have robust conceptual and phonological knowledge, children are better at producing phrases which are high in complexity when they have been practiced more often. Therefore, once the form is in place, we see the frequency driven effect of phonological/motor chunking.

We next conducted a further analysis using the combined data from experiment 1 and experiment 2. This allowed us to explore the effect of age on our variables. For the error data, we found that older children made fewer errors in general. However, for duration data, we did not find any effect of age. These findings tentatively suggest that children do have slightly more robust phonological representations as they get older – since fewer errors were made in general.

7.3 Experiment 3

In experiment 3, we replicated the procedure from experiments 1 and 2 in adults in order to explore the continuation of developmental effects. For this, adults repeated phrases manipulated for phrase frequency, word frequency and phonological complexity. We used duration of production as a measure of fluency and conducted a multilevel linear regression analysis to explore our research questions. The hypotheses were the same as for experiments 1 and 2.

We found a main effect of word frequency where speakers took less time to produce phrases when they contained high frequency words, as well as a main effect of complexity where speakers took less time to produce phrases which were low in complexity. In addition to this, we found an interaction between word frequency and complexity, where speakers took less time to produce complex phrases when they contained high frequency words. These results are in line with our predictions and replicate the findings of experiment 2, thereby providing evidence for adults engaging in phonological/motor chunking.

However, we also found a main effect of phrase frequency where speakers took more time to produce phrases when they were high in frequency and a significant negative interaction between word and phrase frequency. This interaction indicates that that when words are lower in frequency, high frequency phrases take longer to produce than low frequency phrases (the opposite of the effect we would predict for phrase frequency independent of word frequency). However, when words are high in frequency, phrase frequency does not impact speed of production. These latter findings are interesting because they suggest that when phrase frequency is high, the learning of words is impeded. One possible explanation for this finding is that there is competition between the representation of the word and the representation of the phrase (e.g., Sosa & MacFarlane, 2002). This could occur either at the point of learning or the point of production and is worthy of further investigation.

7.4 Experiment 4 and 5

In experiments 4 and 5, we aimed to explore the effect of word frequency and phrase frequency on the accuracy of phoneme production in adults using artificial stimuli. Since adults are being taught novel strings, this experiment allows us to look at less established production knowledge and can be used to compare results to the child data. Furthermore, the results from this experiment can be used to confirm that the results from the previous experiment are not due to systematic frequency-related differences in the meaning of words and phrases. There are some findings which suggest that speakers might have more robust conceptual knowledge of words which have stronger meanings, instead of this effect being driven by production frequency. This idea is most often explained in terms of a spreading activation network (e.g., Dell, 1997), where semantically related words are activated during the conceptualisation stage of production. Harley and MacAndrew (1992) looked at word substitution errors in naturalistic speech and found that high-imageability items have higher activation levels than low-imageability items (low-imageability items were more likely to be replaced with high-imageability items). However, the results from the non-word experiments suggest that the results in this thesis are due entirely to production frequency.

In experiment 4, adults were required to listen to and repeat recordings of novel phrases consisting of non-words. In this experiment, the relative frequencies of the target words and phrases were manipulated via a training stage, where the number of times the speakers repeated target words and phrases varied. Once this training stage was completed, the test phase began. For these experimental trials, the speaker produced each target phrase once and the error rate for these phrases was measured. We hypothesised that speakers would make fewer errors when producing high frequency phrases and/or high frequency words. We also tested the secondary prediction that phrase and word frequency will affect (promote or suppress) each other's effect on accuracy. We found a main effect of word frequency where speakers made fewer errors producing phrases containing high frequency words. There was no main effect of phrase frequency and no significant interaction was seen.

In experiment 5, we used the same stimuli but presented speakers with both the auditory and orthographic representations of the target phrases for repetition. This meant that the speaker was not required to hold the phrase in phonological working memory for the full course of production. The hypotheses were the same as for experiment 4. As in experiment 4, we found a main effect of word frequency, where speakers made fewer errors when producing phrases containing high frequency words. We did not find a main effect of phrase frequency or a significant interaction.

These experiments allowed us to explore the effect of frequency on the entire production process, as well as for articulatory practice in isolation. Taken together, given that we removed the possibility of previous effects being due to systematic frequency-related differences in the meaning of words and phrases, these results suggest that speakers do use both chunked phonological representations as well as optimised muscle activation patterns during phonological/motor chunking, as proposed by Guenther (2016).

7.5 Overall implications

To consider the overall implications of our results, we discuss our findings in terms of possible explanations as to why speakers (and particularly children) find it easier to produce some words/phrases compared to others. In our experiments, we focused on the effects of word frequency, phrase frequency and word complexity (and their interactions) on production. One theory we were particularly interested in is phonological/motor chunking, a process where words/phrases become easier to say after practice. Therefore, we will first summarise previous findings relating to frequency and complexity effects, before focusing on the evidence seen for phonological/motor chunking in our experiments and what this means for the predictions of different models of production.

7.6 Word frequency

Both adults and children have been shown to be more fluent in their production of words which are high in frequency compared to words which are low in frequency. Vitevich and Sommers (2003) found that when adult speakers listened to definitions of words which varied in frequency, more tip of the tongue states were elicited for target words with low frequency, compared to target words with high frequency. Harley and MacAndrew (2001) show that more phonological errors occur in naturalistic speech for less frequent words, though more complex (longer) words were also found to elicit more errors. This effect is also well substantiated in children. Sosa and Stoel-Gammon (2012) show during an elicited imitation task, 2;0-2;5-year-old children's productions of high frequency words were less variable than those of low frequency words, though more complex words (containing later developing consonants and syllable structures) were also produced more variably.

Different models of word production have sought to explain this frequency effect. Although there are many diverse models of speech production (e.g., Garrett, 1975; Caramazza, 1997), most models contend that the production of words requires 2 stages. Generally, the most accepted account of lexical access is Levelt's (1989; 2001) two-stage lexical access model. In this model, the speaker first specifies a concept, which leads to activation of the linguistic form (lemma). The lemma also represents a word's syntactic and semantic information. Here, other lemmas with similar contexts are also activated. There is competition between syntactically and semantically related lemmas while lemma selection occurs. Once the target lemma has been selected, the lexeme (phonological form of the word) is realised before the final stage of articulation, which involves motor execution and production of the word.

In this model, word frequency effects are attributed to the second (phonological) stage of lexical access. The speed of access to phonological forms is constrained by word frequency, so that speakers are faster to access the phonological forms of words which are more frequent, compared to the phonological forms of words which are less frequent (see Levelt, 2001 for description). However, there is some evidence to suggest that word frequency does have an effect at the semantic level (Navarette, Basagni, Alorio & Costa, 2006). This model of lexical access does not explicitly account for phrase frequency effects.

In our experiments, we found an effect of word frequency, where both adults and children made less errors/took less time to produce phrases when they contained words which were high in frequency. We argue, that at least for 3- and 4-year-olds, this word frequency effect is seen both during conceptualisation and formulation stages.

7.7 Phrase frequency

This effect is also well substantiated in phrases. Bannard and Matthews (2008) tested 2- and 3-year-old children's knowledge of frequent multiword sequences via a repetition task using pairs of sequences that were identical except for the final word. Frequently occurring sequences (e.g., *a piece of cheese*) were matched to infrequent sequences (e.g., a piece of food), where the final words (e.g., *cheese, food*) and final bigrams (*of cheese, of food*) were matched for frequency. Here, children were found to repeat the first 3 words of the high-frequency word sequence combinations more quickly and more accurately than the low-frequency combinations. This finding has since been replicated using different stimuli with older children (Kueser & Leonard, 2020) and using a similar methodology to Bannard and Matthews (2008), Arnon and Cohen Priva (2013) further found that adults showed reduced phonetic duration for high frequency phrases compared to low frequency phrases when repeating high (*don't have to worry*) vs low (*don't have to wait*) frequency phrases of the same syntactic type, where word, bigram and trigram frequency were held constant.

In experiment 3 (adult real language sentence repetition task) we found a main effect of phrase frequency where speakers took more time to produce phrases which were high in frequency. This is the opposite of what was predicted. We also found an interaction between phrase frequency and word frequency so that speakers took more time to produce high frequency phrases than low frequency phrases, when they contained words which were low in frequency. One possible explanation for this finding is that there is competition between the representation of the word and of the phrase (see Sosa & MacFarlane, 2002). If phonological/motor chunking is driven by word frequency, as shown in the previous experiments, it might be that the retrieval of less frequent words is impeded even more so when these words are contained in frequent phrases.

However, no effect of phrase frequency was seen in our other experiments. The absence of phrase frequency effects in the child and non-word data is surprising given the abundance of work showing that phrase frequency does influence production fluency in both children and adults. A possible explanation for this is that much of the previous literature focussing on phrase frequency effects uses longer phrases. For example, Bannard and Matthews (2008) tested 2- and 3-year-old children's knowledge of frequent multiword sequences via a repetition task using pairs of 4-word sequences that were identical except for the final word and found that children were faster and made less errors when producing high frequency phrases. Similarly, Arnon and Priva (2013) used 4-word sequences to explore phonetic duration in adults, where it was shown that phonetic duration is reduced for these high frequency phrases. Later, Arnon and Cohen Priva (2014) replicated this finding using trigrams.

Our study differs to that done previously in that it looks at two-word phrases in the repetition tasks. The reason for this design decision was that we assumed 2 syllables are an upper bound on the size of a chunk a 3-year-old could plausibly form a phonological chunk form. The relevance of the length of the phrase is that this introduces memory demands that do not exist with our two-word phrases. It could then be that phrase frequency effects are driven by memory. One account of the role of working memory in recall for phrases that might account for such effects is provided by Jacobs, Dell and Bannard (2017). They propose that phrase frequency effects on recall are driven by redintegration wherein long-term memory representations support short term memory. It is possible that redintegration only has an impact when the working memory demands are greater (when the phrase is longer). This needs further investigation.

7.8 Frequency and complexity

We know that frequency has an effect at both the word and phrase level, where frequent words/phrases are produced more fluently than infrequent words/phrases. However, frequency also plays a role in how difficult speakers find different sounds and words to produce, where practiced complex sequences are produced more fluently than infrequent complex sequences (e.g., Sosa & Stoel-Gammon, 2012). One explanation for this, where speakers become more proficient in their production for complex words that have been practiced more often, is called phonological/motor chunking.

Assuming that errors do come from the formulation stage, frequency-related effects in adult production are usually conceptualised, even if only implicitly, as an effect of phonological/motor sequence learning, with words that are difficult to articulate when first encountered becoming easier to say due to practice. One model of production which directly considers this is the GODIVA model (Guenther, 2016). The idea is that frequent sounds are represented together as a single unit. This means that the production system requires a reduced working memory load to access the target form. In addition to this, sequences are produced more fluently because the articulatory program has been practiced more often in comparison to less frequent sequences.

The GODIVA model assumes that the production system stores frequently occurring sub sequences as cohesive ‘chunks’ which reduce phonological working memory load and improve motor performance. For non-optimised targets, once the target sequence is loaded, the structure and phonetic content are chosen based on activity gradient. Meanwhile, the sequential structure buffer also projects to the initiation map, which becomes active when the current context for the current sound is recognised. At the same time, the phonological content buffer signals the upcoming phonological items to the speech sound map, which chooses the best motor program for producing these items. For non-optimised targets, the speech sound map activates each phoneme in sequence. Finally, articulation occurs when the associated initiation map node is activated.

However, for optimised targets, working memory buffers in phonological content buffer (pIFS) and sequential structure buffer (preSMA) contain cluster sized SSCs to reduce the number of items that are stored in working memory. SSCs refer to the way a syllable is broken into 3 sub-syllabic constituents. These are the onset (one or more consecutive consonants at the beginning of the syllable), nucleus (vowel, diphthong or sonorant consonant) and coda (one or more consecutive consonants at the end of the syllable). For optimised targets, the gestures for these SSCs are mediated by the basal ganglia loop instead of the sequential structure buffer (preSMA), which also activates gestural motor programs in the initiation map cells in the initiation map (SMA). Meanwhile, the speech sound map activates the syllable/word as a chunk and subcortical loops in the cerebellum coordinate and coarticulate the individual motor gestures. Finally, articulation occurs when the associated initiation map node is activated.

This model contends that word frequency has an effect at both the word retrieval and phonological level. During the first stage, speakers take less time to retrieve words containing chunked consonant clusters. During the second phonological stage, there is a processing advantage for whole words which are high in frequency, where speakers make less errors for frequent words since they are stored as chunked motor programs. Although this model does not yet explicitly consider the effect of phrase frequency, the model also does not explicitly consider words, but motor sequences. Therefore, it could be argued that phrase frequency effects would just be the result of chunks that cross word boundaries.

One study which provides direct evidence for phonological/motoric chunking was conducted by Segawa et al. (2015), who explored the trade-off between phonological complexity and frequency using phonotactic probability as a measure of complexity. Adults were presented with phototactically legal and illegal triconsonantal initial and final consonant clusters contained in mono-syllabic non-words. It was found that speakers had a significantly lower error rate after practice for the illegal sequences but did not differ in their repetition of legal syllables. The authors took this latter finding as evidence for learned motor programs - given that performance improvement was specific to the stimuli encountered, production could not be a result of the learning of phonological rules; instead improvement is a consequence of phonological/motor learning.

7.9 Observed word frequency by complexity interaction

We predicted that if phonological/motor chunking is taking place, we would see an interaction between phrase/word frequency and complexity, where complex words/phrases would be produced more fluently when they were more frequent, compared to complex words/phrases which are less frequent.

Based on this assertion, the account best supported by our results from experiment 1 and 2 suggests that children do engage in phonological/motor chunking, but only once they have robust knowledge of the target lemma and/or the process used during lemma retrieval. We found that children made fewer errors when producing low complexity phrases but only when they contained high frequency words, and that children struggled to produce phrases which were low in frequency regardless of complexity.

These findings suggest that there is an effect of word frequency at the conceptualisation level, where children do not seem to have strong representations for words which are not frequent. Children are only able to fully retrieve lemmas for frequent words and then show an advantage in producing the frequent words which are not complex. This conclusion would be incompatible with an assumption of the two-stage model proposed by Levelt et al. (1999) who argue that frequency effects are only seen during the formulation stage. However, this model is only intended as an account of the production process for adult speakers with robust conceptual and phonological knowledge. It does not account for children. It is possible that this frequency effect seen during the conceptualisation stage in our experiments is only apparent in children at this age/during development. It seems likely that once speakers have full conceptual knowledge (as adults do), this frequency advantage is no longer seen.

Further support for this account is provided by a duration analysis of our data. For phrases which were produced correctly (so for phrases in which children do have robust conceptual and phonological knowledge of the lemma) we found that, overall, children took less time to produce complex phrases when they contained high frequency words. This finding does provide evidence of phonological/motor chunking (e.g., Guenther, 2016) which refers to the idea that the production system stores frequently occurring sub sequences as cohesive ‘chunks’ in phonological working memory. Our findings, which show that complex phrases that have been practiced more often are produced more quickly, indicate that phonological/motor chunking is taking place. Children take longer to produce complex phrases which are low in word frequency because these phrases must be formed step by step. Furthermore, the observed phonological/motor chunking appears to be driven by word frequency, given that no main effect of phrase frequency was seen for these children.

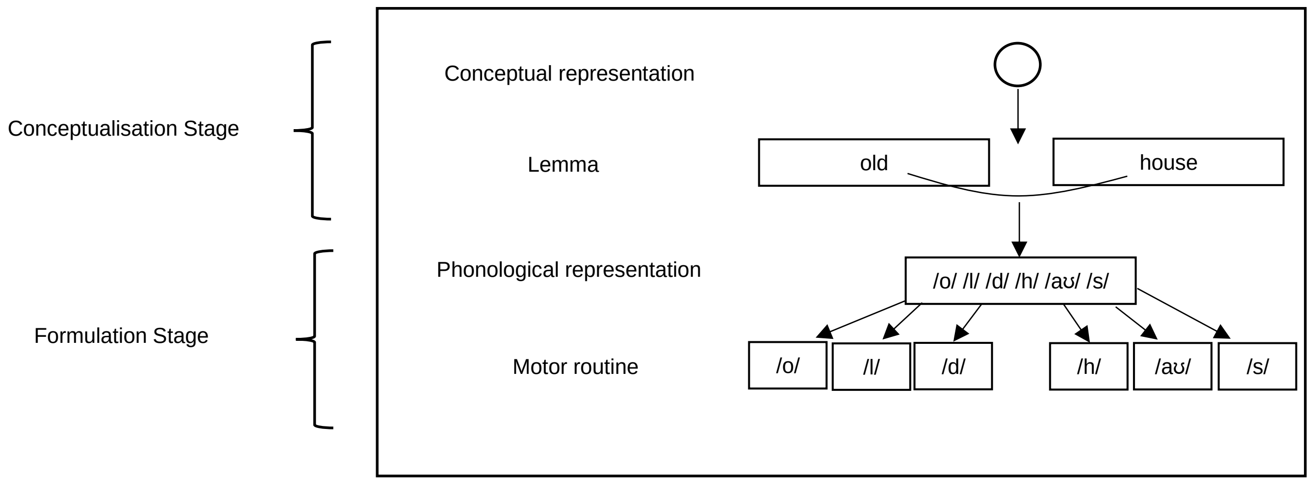
This word frequency by complexity interaction was also seen in adults during experiment 3, where speakers took less time to produce complex phrases when they also contained frequent words. This further supports the idea that phonological/motor chunking is driven by word frequency, given that adults showed a production advantage for complex words more quickly when they had been practiced more often.

The idea that phonological/motor chunking is driven by word frequency is further supported by the results from the non-word experiments, where we found that adults were more likely to produce phrases correctly when they were high in word frequency, while no effect of phrase frequency was seen. These results also show that word frequency has an effect at the formulation stage of production which includes both the phonological working memory level (where motor programs are stored) and the articulatory level (where practice of gestures leads to faster production). Segawa et al. (2019) argue that stored frequent consonant clusters drive frequency effects at the phonological working memory stage, while practice of entire words leads to optimised muscle activation patterns during articulation, although this is not clear from our results.

7.10 <u>Informal account of production</u>

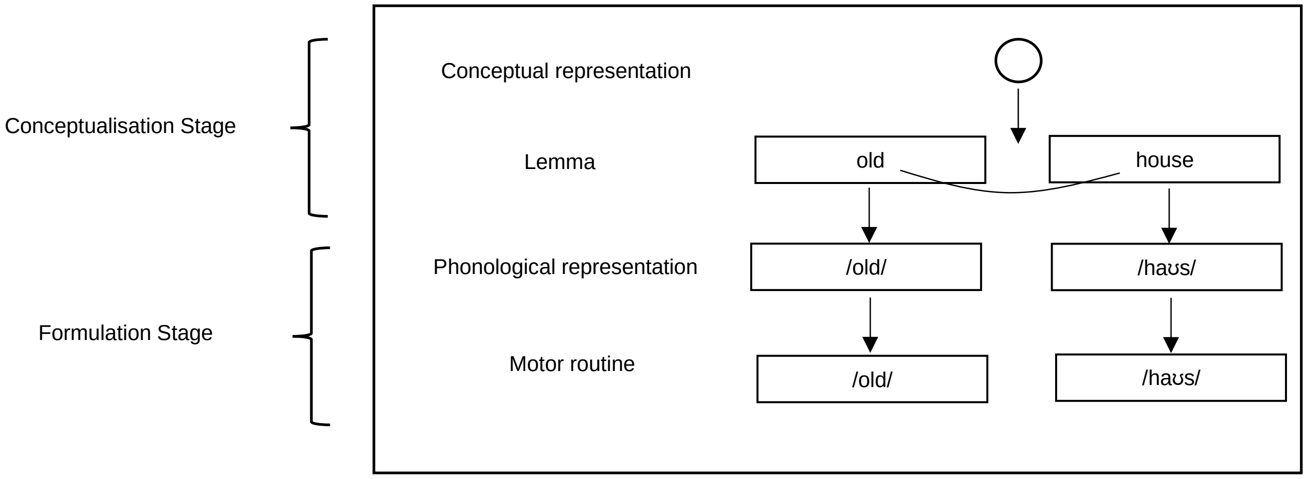
In the following we propose an informal account of production which accounts for our findings (see Figure 7-1). As is the case for many models of production, we contend that there is a semantically specified representation followed by a phonologically specified representation. This system is organised by frequency, with more frequent words/sounds being more easily accessed. We argue that, at least for young children who do not have robust conceptual and phonological knowledge, there are frequency effects at both the conceptualisation and formulation stage.

Figure 7-1: <u>Informal Model of Production</u>



To produce a word, the speaker is first required to retrieve the lemma. Lemmas for words which are more frequent are more easily accessed. Next, during the formulation stage, the phonological information for the target word is realised. There is a processing advantage for frequent words during this stage, at least in terms of words for which the speaker has robust conceptual and phonological knowledge. This advantage is due to phonological/motor chunking, where speakers become more proficient in their production due to practice (see Figure 7-2).

Figure 7-2: <u>Informal Model of Production using Chunking</u>



It is possible that chunking takes place in the form of chunked sequences of phonemes (phonological) for words which are more frequent, which allows for more rapid retrieval of phonological information. It is also possible that chunking takes place in the form of a learned motor routine (motoric), where articulatory movements are more rapid for more frequent words due to practice. It is not possible to distinguish between these two accounts, but we assume that a chunking process is driving the observed word frequency effect during the formulation stage, seen in our experiments.

To produce a novel word, there will be no lemma/representations available. In this case, the speaker must rely on knowledge of phonemes. This means that production is more difficult and takes more time.

The results in our experiments suggest that phonological/motor chunking occurs for word but not phrase frequency. This suggests that chunking at this level does not play a role in observed phrase frequency effects. However, we did observe competition between word frequency and phrase frequency when measuring duration of production in adult speakers for real language. It is most likely that this competition occurs during the conceptualisation stage of production. To produce a phrase, the speakers must first retrieve the phrasal representations consisting of the constituent lemmas, which are held together by long term memory associations. If we accept that frequency effects are observed during this conceptualisation stage, it is possible that the retrieval of low frequency lemmas is impeded when they are also contained in high frequency phrases.

Words which occur together frequently during production have stronger long term memory associations with one another compared to words which do not occur together as frequently. Therefore, it is possible that activation of the initial adjective lemma leads to activation of the lemmas for a range of nouns for which there are memory associations. This could lead to competition between the activation of different nouns and the activation of the overall target phrase. For adults, it is also possible that a more abstract representation (e.g., adjective noun or NP) is activated and that this competes with activation of the individual low frequency lemmas for the constituent words.

7.11 Limitations

A possible limitation of these experiments is that our measure of frequency is based on adult speech. We assume that the frequency of words/phrases in child speech mirrors that of adults, leading to the prediction that frequency effects in child speech (and the associated phonological/motor chunking) would mirror those in adult speech. However, if we accept our interpretation of the error data, where the difficulty children face in producing low frequency words is due to an earlier stage than articulation, it is also possible that this is as much a comprehension as a production problem. By this account, it is input frequency which is driving the effect seen at the conceptualisation level so that children only have robust conceptual knowledge of words that they have heard often. In this case, the effect is more direct, so that children have immature knowledge of infrequent words not because of their own experience of production, but simply because they have not heard these words as often as the frequent words.

It is also possible that the frequency effect observed is due to the comprehension part of the repetition task. We suggest that to take part in a sentence repetition task, the speaker is first required to retain and recall the target sentence. The episodic buffer supports this short-term retention during sentence repetition by integrating representations of words/phrases (from working memory and long-term memory); where there are strong representations of words and phrases, retention will be well supported. Chunking/frequency effects are rooted in the episodic buffer before the speaker uses the articulatory system to produce the target sentence. By this model, the speaker is required to first comprehend and then regenerate the phrase. Therefore, it is possible that the observed frequency effect is observed during initial comprehension (before the speaker starts to regenerate the phrase).

Finally, it is possible that the stimuli that were used in experiment 1, 2 and 3 are simply too low in frequency to provide data relevant to the key questions of interest. This limitation possibly explains the lack of a phrase frequency effect seen in experiment 1 and 2 (3- and 4-year-old real language experiments), although it is not clear why this effect was still not observed in experiment 4 and 5 (non-word experiments).

7.12 Future research

Our results provide evidence for both children and adults using phonological/motor chunking for 2-word phrases, a process which appears to be driven by word frequency. We chose to use 2-word phrases based on the assumption that 2 syllables is an upper bound on the size of chunk a 3-year-old could plausibly form. However, it would be interesting to use a similar elicited imitation task with 2 syllable words to see whether the word has a special status – whether phonological/motor chunking is being driven by syllable frequency and would be seen for sub-word units. If chunking is being driven by word frequency then we would expect to see the same word frequency by complexity interaction, where speakers make less errors for complex words when they are more frequent. However, if chunking is being driven by syllable frequency, we would expect to see the same directional interaction but for syllable frequency and complexity.

Furthermore, it would be useful to extend experiment 4 and 5 to include phonological complexity – an interaction effect between word frequency and complexity would further confirm that the observed word frequency effect is due to phonological/motor chunking. To do this, the non-words could be manipulated in a similar way as in our real language experiments (for example using WCM as a measure of complexity) where frequency would be manipulated in the same way so that all 3 factors are fully crossed.

Our informal model considers the possibility that phrase frequency effects are driven by redintegration, but only when working memory demands are greater (when the phrase is longer). If this is the case, we would expect to see a production advantage for phrase frequency but no interaction between phrase frequency and complexity - or perhaps an interaction in the other direction, where speakers have a production advantage for high frequency phrases when they are less complex. More research is required to explore this.

Finally, our experiments are the first to cross factors of word and phrase frequency. We observed an interaction between word frequency and phrase frequency which we took as evidence of competition between representations. We suggest that, if we accept frequency effects are observed during the conceptualisation stage, it is possible that the retrieval of low frequency lemmas is impeded when they are also contained in high frequency phrases. Therefore, work is required to explore this interaction further.

7.13 Conclusion

This thesis aimed to explore the effect of phrase frequency, word frequency and complexity, as well as their interactions, on the production fluency of both children and adults. We aimed to explore whether the previous findings for independent effects of phrase frequency, word frequency and word complexity on phrase production would hold when using phrases that are simultaneously manipulated for all three factors, and second to understand how the effects of these properties interact.

We found that children and adults were more fluent in their production of phrases when they contained more frequent words. However, an effect of phrase frequency was only seen in the adult real language data, where phrases took longer to produce when they were more frequent. This is the opposite of what was predicted. We were specifically interested in whether children, like adults, engage in phonological/motor chunking. Evidence of this would be seen through an interaction between phrase/word frequency and complexity, where more complex words/phrases would be produced more fluently when they were also more frequent, compared to complex infrequent words/phrases. This is because words become easier to say after practice as a result of motor learning. Taken together, our findings suggest that children do seem to use phonological/motor chunking, but only when they have robust knowledge of the conceptual and phonological information contained in the target phrase. This is further supported by the complexity by word frequency interaction seen in the same experiment in adults, while phonological/motor chunking was driven by word frequency in adults when using artificial stimuli.

Furthermore, our real language experiments were the first to cross factors of word and phrase frequency. We aimed to explore whether these effects are independent and, if not, whether they are complementary or if the accessing of representations at different levels might lead to competition. We observed an interaction between word frequency and phrase frequency in the adult study which suggests that production mechanisms underlying word frequency and phrase frequency compete during production. We found no evidence of phrase frequency effects in our other experiments. We take this as evidence that phrase frequency does not drive phonological/motor chunking. Instead, it is possible that the phrase frequency effects seen in other research are driven by working memory, though more work is required to explore this.

References

Alloway, T. P., Gathercole, S. E., Willis, C. & Adams, A. M. (2004). A structural analysis of working memory and related cognitive skills in young children. *Journal of Experimental Child Psychology*, *87*. 85–106.

Anderson, J. D. (2007). Phonological Neighborhood and Word Frequency Effects in the Stuttered Disfluencies of Children Who Stutter. *Speech Lang Hear Res, 50*(1). 229-247.

Arnon, I & Priva, C. U. (2013). More than Words: The Effect of Multi-word Frequency and Constituency on Phonetic Duration. *Language and Speech. 56*(3). 349–371.

Arnon, I & Priva, C. U. (2014). Time and again: The changing effect of word and multiword frequency on phonetic duration for highly frequent sequences. *The Mental Lexicon. 9*(3). 349-371.

Arnon, I., & Snider, N. (2010). More than words: Frequency effects for multi-word phrases. *Journal of Memory and Language, 62*. 67–82.

Baddeley, A. (2000). The episodic buffer: a new component of working memory? *Trends in Cognitive Sciences, 4*(11). 417–423.

Baddeley, A. D., & Hitch, G. J. (1974), Working memory. In G. Bower (ed.), *The Psychology of Learning and Motivation* (New York, NY: Academic Press), pp. 47–87.

Bannard (2006): Bannard, C. (2006). *Acquiring phrasal lexicons from corpora* (Doctoral dissertation, University of Edinburgh).

Bannard, C & Matthews, D. (2008) Stored Word Sequences in Language Learning: The Effect of Familiarity on Children’s Repetition of Four-Word Combinations. *Psychological Science, 19*(3). 241 – 248.

Beattie, G. W., & Butterworth, B. L. (1979). Contextual probability and word frequency as determinants of pauses and errors in spontaneous speech. *Language and Speech, 22.* 201–211.

Bell, A., Jurafsky, D., Fosler-Lussier, E., Girand, C., Gregory, M., & Gildea, D. (2003). Effects of disfluencies, predictability, and utterance position on word form variation in English conversation. *The Journal of the Acoustical Society of America, 113.* 1001–1024.

Bock, K., & Levelt, W.J.M. (1994). Language production: grammatical encoding. In M.A. Gernsbacher (Ed.), Handbook of psycholinguistics (pp. 945–984). San Diego, CA: Academic Press.

Brown, R. (1973). *A first language: the early stages*. Cambridge, MA: Harvard University Press.

Bybee, J. (2000). The phonology of the lexicon. In M. Barlow & S. Kemmer (Eds.), *Usage-based models of language* (pp. 65–85). Stanford, CA: CSLI.

Bybee (2006). From Usage to Grammar: The Mind's Response to Repetition. *Linguistic Society of America, 82*(4). 711-733.

Bybee, J., & Scheibman, J. (1999). The effect of usage on degree of constituency: The reduction of *don't* in American English. *Linguistics, 37.* 575–596.

Caramazza, A. (1997). How many levels of processing are there in lexical access? *Cognitive Neuropsychology, 14*(1). 177-208.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and psychological measurement*, *20*(1). 37-46.

Conti-Ramsden, G., Botting, N., & Faragher, B. (2001). Psycholinguistic markers for specific language impairment (SLI). *Journal of Child Psychology & Psychiatry & Allied Disciplines, 42*(6). 741–748.

Dell, G. S. (1990). Effects of frequency and vocabulary type on phonological speech errors. *Language and Cognitive Processes, 5*. 313–49.

Dell, G. S. & O'Seaghdha, P. G. (1992). Stages of lexical access in language production. *Cognition, 42.* 287-314.

Dell, G. S., Schwartz, M. F., Martin, N., Saffran, E. M., & Gagnon, D. A. (1997). Lexical access in aphasic and nonaphasic speakers. *Psychological Review*, *104*. 801–838.

Dell, G. S., & Reich, P. A. (1981). Stages in sentence production: An analysis of speech error data. *Journal of Verbal Learning and Verbal Behavior*, *20*. 611–629.

Dworzynski, K. & Howell, P. (2004). Predicting stuttering from phonetic complexity in German. *Journal of Fluency Disorders*, *29*. 149–173.

Ellis, N. (2002). Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition, 24*(2). 143-188.

Finocchiaro, C. & Caramazza, A. (2006). The production of pronominal clitics: Implications for theories of lexical access. *Language and Cognitive Processes*, 21 (1/2/3). 141-180.

Garrett, M. F. (1975) The analysis of sentence production. *The psychology of learning and motivation, 9.* 133-177.

Goldrick, M. & Larson, M. (2008). Phonotactic probability influences speech production. *Cognition*, *107*. 1155–1164.

Gobet, F., Lane, P. C. R., Croker, S., Cheng, P. C-H., Jones, G., Oliver, I. & Pine, J. M. (2001). Chunking mechanisms in human learning. *Trends in Cognitive Sciences, 5.* 236-243.

Green, P. & MacLeod, C. J. 2016. SIMR: An R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution, 7*(4). 493-498.

Guenther, F. H. (2016). *Neural control of speech*. Cambridge, MA: MIT Press.

Guenther, F. H., Ghosh, S. S., & Tourville, J. A. (2006). Neural modeling and imaging of the cortical interactions underlying syllable production. *Brain & Language, 96*. 280–301.

Hayes, B (2011). *Introductory Phonology*. Malden, MA: Wiley-Blackwell.

Harley, T. A. & Brown, H. E. (1998). What causes a tip-of-the-tongue state? Evidence for lexical neighbourhood effects in speech production. *British Journal of Psychology, 89*. 151-174.

Harley, T. A., & MacAndrew, S. B. G. (1992). Modelling paraphasias in normal and aphasic speech. *Proceedings of the 14th Annual Conference of the Cognitive Science Society.* 378–383.

Harley, T. A. & MacAndrew, S. B. G. (2001). Constraints Upon Word Substitution Speech Errors. *Journal of Psycholinguistic Research, 30*(4). 395-418.

Hendrix, P., Baeyen, H. & Bolger, P. (2017). Distinct ERP Signatures of Word Frequency, Phrase Frequency, and Prototypicality in Speech Production. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 43*(1). 128-149.

Hickok, G. (2012). The cortical organization of speech processing: Feedback control and predictive coding the context of a dual-stream model. *Journal of Communication Disorders, 45*. 393–402.

Hickok, G. & Poeppel, D. (2004). Dorsal and ventral streams: A framework for understanding aspects of the functional anatomy of language. *Cognition*, 92. 67–99.

Hickok, G. & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Publishing Group, 8*. 393-402.

Howell. P. & Au-Yeung, J. (2007). Phonetic complexity and stuttering in Spanish. *Clinical Linguistics & Phonetics, 21*(2). 111–127.

Ingram, D. (2002). The measurement of whole-word productions. *Journal Child Language*, *29*. 713-733.

Jacobs, C. L, Dell, G. S & Bannard, C. (2017). Phrase frequency effects in free recall: Evidence for redintegration. *Journal of Memory and Language, 97*. 1-16.

Jacobs, C. L., Dell, G. S., Benjamin, A. S., & Bannard, C. (2016). Part and whole linguistic experience affect recognition memory for multiword sequences. *Journal of Memory and Language, 87.* 38–58.

Jakielski, K.J. (1998). Motor organization in the acquisition of consonant clusters. *Dissertation Abstracts International*, 59(06B), 2716.

Janssen, N & Barber, H. A. (2012). Phrase Frequency Effects in Language Production. *PLoS ONE, 7*(3). 1-11.

Janssen, N., Bi, Y. & Caramazza, N. (2008). A tale of two frequencies: Determining the speed of lexical access for Mandarin Chinese and English compounds. *Language and Cognitive Processes, 23*(7/8). 1191-1223.

Jescheniak, J. D., & Levelt, W. J. M. (1994). Word frequency effects in speech production: Retrieval of syntactic information and of phonological form. *Journal of Experimental Psychology: Learning, Memory and Cognition, 20*. 824-843.

Jones, S. D. (2020). Accuracy and variability in early spontaneous word production: The effects of age, frequency and neighbourhood density. *First Language*, *40*(2). 128–150.

Jusczyk, P. Q., Luce, P. A. & Charles-Luce, J. (1994). Infant's sensitivity to phonotactic patterns in the native Language. *Journal of Memory and Language, 33*(5). 630-645.

Kittredge, A. K., Dell, G.S., Verkuilen, J & Schwartz, M. F. (2008). Where is the effect of frequency in word production? Insights from aphasic picture-naming errors. *Cognitive Neuropscyhology*, 25(4). 463–492

Knobel, M., Finkbeiner, M., & Caramazza, A. (2008). The many places of frequency: Evidence for a novel locus of the lexical frequency effect in word production. *Cognitive Neuropsychology, 25.* 256–286.

Kueser, J. B., & Leonard (2020). The Effects of Frequency and Predictability on Repetition in Children With Developmental Language Disorder. Journal of Speech, Language, and Hearing Research, *63*. 1165–1180.

LaSalle, L. R., & Wolk, L. (2011). Stuttering, cluttering and phonological complexity: Case studies. *Journal of Fluency Disorders*, *36*. 285–289.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge. MA: MIT Press.

Levelt, J. M., Roelofs, A. & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences, 22*. 1-75.

Levelt, W J. M., Schriefers, H., Vorberg, D, Meyer, A. S, Pechmann, T, & Havinga, J. (1991). The time course of lexical access in speech production: A study of picture naming. *Psychological Review, 98*. 122-142

Levelt, W. J. M., & Meyer, A. S. (2000). Word for word: Multiple lexical access in speech production. *European Journal of Cognitive Psychology, 12*(4). 433-452

Lindblom, B & Maddieson, I. (1988). *Phonetic Universals in Consonant Systems*. In Li, C. and Hyman, L. M. (eds.), Language, Speech and Mind, 62-78. London: Routledge.

Lombardi, L. and Potter, M. C. (1992). The regeneration of syntax in short term memory. *Journal of Memory and Language*, *31*(6). 713–733.

Maddieson, I. (1984). *Patterns of Sounds*. Cambridge University Press, Cambridge.

Maddieson, I. (2006). Correlating phonological complexity: data and validation. *Linguistic Typology*, *10*, 108-125.

Maddieson, I. (2009). Calculating phonological complexity. *Approaches to Phonological Complexity, Phonology and Phonetic Series*, *16.* 85-110.

Moreland, M. (2011). The Status of the Post-Initial Jod: Evidence from British English Wordlikeness Judgements. University of Reading.

Munson, B. & Solomon, N. P. (2004). The effect of phonological neighborhood density on vowel articulation. *J Speech Lang Hear Res*, *47*(5). 1048–1058.

Navarette, E., Basagni, B., Alario, F.X. & Costa, A. (2006). Does word frequency affect lexical selection in speech production? *The Quarterly Journal of Experimental Psychology, 59*(10). 1681-1690.

Needleman, S. B. & Wunsch, C. D. (1970). A General Method Applicable to the Search for Similarities in the Amino Acid Sequence of Two Proteins. *Journal of Molecular Biology, 48*(3). 443-453.

Newman and German (2005). Life Span Effects of Lexical Factors on Oral Naming*. Language and Speech, 48*(2). 123-156*.*

Newman, R. S. & German, D. J. (2002). Effects of lexical factors on lexical access among typical language-learning children and children with word-finding difficulties. *Language and Speech, 45*(3). 285-317.

Nickels, L. (1995). Getting it right? Using aphasic naming errors to evaluate theoretical models of spoken word recognition. *Language and Cognitive Processes,* 10. 13–45.

Oldfield, R. C. & Wingfield, A. (1965). Response latencies in naming objects. *Quarterly Journal of Experimental Psychology, 17*. 273–81.

Ota, M., & Green, S. J. (2013). Input frequency and lexical variability in phonological development: a survival analysis of word-initial cluster production. *Child Lang, 40*. 539–566.

Paul Boersma & David Weenink (2018): Praat: doing phonetics by computer [Computer program]. Version 6.0.39 http://www.praat.org/

Peterson, R. R., & Savoy, P. (1998). Lexical selection and phonological encoding during language production: Evidence for cascaded processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24(*3). 539–557.

Pinker, S. & Ullman, M. T. (2002). The past and future or the past tense debate. *Trends in Cognitive Sciences*, *6*(11). 456–463.

Potter, M. C., & Lombardi, L. (1990). Regeneration in the short-term recall of sentences. *Journal of Memory and Language, 29*. 633-654.

Rapp, B., & Goldrick, M. (2000). Discreteness and interactivity in spoken word production. *Psychological Review*, *107*. 460–499.

Reali, F., & Christiansen, M. H. (2007). Processing of relative clauses is made easier by frequency of occurrence. *Journal of Memory and Language, 57*(1). 1-23.

Riches, N. G. (2012). Sentence repetition in children with specific language impairment: An investigation of underlying mechanisms. International Journal of Language & Communication Disorders. 47(5). 499-510.

Santesteban, M., Costa, A., Pontin, S. & Navarrete, E. (2006). The effect of word-frequency on lexical selection in speech production: Evidence from semantic homogeneous naming contexts. *Cognitiva, 18*(1). 75-84.

Sasisekaran, J., Smith, A., Sadagopan, N. & Weber-Fox, C. (2010). Nonword repetition in children and adults: effects on movement coordination. *Developmental Science, 13*(3). 521–532.

Schriefers, H. (1993). Syntactic processes in the production of noun phrases. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*. 841–850

Schwartz, R. & Leonard, L. (1982). Do children pick and choose? An examination of phonological selection and avoidance in early lexical acquisition. *Journal of Child Language, 9*. 319–36.

Segawa, J., Masapollo, M., Tong, M., Smith, D. J., & Guenther, F. H. (2019). Chunking of phonological units in speech sequencing. *Brain and Language*, *195*. 1-6.

Segawa, J., Tourville, J., Beal, D., & Guenther, F. H. (2015). The neural correlates of speech motor sequence learning. *Journal of Cognitive Neuroscience, 27*(4). 1-13.

Shantz, K. (2016). Phrase frequency, proficiency and grammaticality interact in non-native processing: Implications for theories of SLA. *Second Language Research, 33*(1). 91-118.

Shao, Z., van Paridon. K & Poletiek, F. (2019). Effects of Phrase and Word Frequencies in Noun Phrase Production. *American Psychological Association*, *45*(1). 147–165.

Siyanova-Chanturia, A., Conklin, K., & Schmitt, N. (2011). Adding more fuel to the fire: An eye-tracking study of idiom processing by native and nonnative speakers. *Second Language Research, 27.* 251–272.

Siyanova-Chanturia, A., Conklin, K., & van Heuven, W. J. B. (2011). Seeing a phrase "time and again" matters: The role of phrasal frequency in the processing of multiword sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37.* 776–784.

Siyanova-Chanturia, A., Conklin, K., Caffarra, S., Kaan, E & van Heuven, W. J. B. (2017). Representation and processing of multi-word expressions in the brain. *Brain and Language, 175.* 111–122.

Sosa, A. V, MacFarlane, J. (2002). Evidence for frequency-based constituents in the mental lexicon: Collocations involving the word of. *Brain and Language, 83*. 227–236.

Sosa, A. V., & Stoel-Gammon, C. (2012). Lexical and phonological effects in early word production. *Journal of Speech, Language, and Hearing Research, 55*. 596-608.

Stemberger, J. P. (1984). Structural errors in normal and agrammatic speech. *Cognitive Neuropsychology, 1.* 281–313.

Stemberger, J. P., & MacWhinney, B. (1986). Frequency and the lexical storage of regularly inflected forms. *Memory & Cognition, 14*(1). 7-26.

Stoel-Gammon, C. (2010). The word complexity measure: Description and application to developmental phonology and disorders. *Clinical Linguistics & Phonetics*, 24. 271–282.

Storkel, H. L. (2001). Learning new words: phonotactic probability in language development. Phonotactic probability in language development. *Journal of Speech, Language, and Hearing Research, 44.* 1321-1337.

Strijkers, K., Costa, A. & Thierry, G. (2010). Tracking Lexical Access in speech production: Electrophysiological correlates of word frequency and cognate effects. *Cerebral Cortex*, *20*(4). 912-928.

Szmaragdand, C. & Leckie, G. (2013) *Introduction to multilevel modelling – R practical*. LEMMA VLE Module, 11, pp. 1-38. Retrieved from http://www.bristol.ac.uk/cmm/learning/course.html

Tremblay, A., & Baayen, R. H. (2010). Holistic processing of regular four-word sequences: A behavioral and ERP study of the effects of structure, frequency, and probability on immediate free recall. In D. Wood (Ed.), *Perspectives on formulaic language: Acquisition and communication* (pp. 151–173). London, UK: Continuum International.

Tremblay, A., Derwing, B., Libben, G., & Westbury, C. (2011). Processing advantages of lexical bundles: Evidence from self-paced reading and sentence recall tasks. *Language Learning, 61.* 569–613.

Underwood, G., Schmitt, N., & Galpin, A. (2004). The eyes have it: An eye movement study into the processing*. Formulaic Sequences: Acquisition, Processing and Use*. 153-172.

Van Lancker Sidtis, D. (2012). Formulaic Language and Language Disorders. *Annual Review of Applied Linguistics. 32*. 62-80.

Vitevich, M. S. & Sommers, M. S. (2003). The facilitative influence of phonological similarity and neighborhood frequency in speech production in younger and older adults. *Memory & Cognition*, *31*(4). 491–504.

Vitevitch, M. S. (2002). The influence of phonological similarity neighborhoods on speech production. *Journal of Experimental Psychology: Learning, Memory & Cognition*, *28*. 735–747.

Vitevitch, M. S., & Luce, P. A. (1998). When words compete: Levels of processing in spoken word perception. *Psychological Science*, *9*. 325-329.

Vitevitch, M. S., & Luce, P. A. (2004). A Web-based interface to calculate phonotactic probability for words and nonwords in English. *Behavior Research Methods, Instruments, & Computers, 36(3).* 481-487.

Wolk, L. & LaSalle, L. R. (2015). Phonological complexity in school-aged children who stutter and exhibit a language disorder. *Journal of Fluency Disorders, 43*. 40–53.

Zaumuner, T. S. (2003). Input-based phonological acquisition. *Input-based Phonological Acquisition*, *62*. 1-106.

Zaumuner, T. S., Gerken, L. & Hammond, M. (2004). Phonotactic probabilities in young children's speech production. *J. Child Lang*, *31*. 515–536

Zimmerer, V. C., Newman, L., Thomson, R., Coleman, M. & Varley, R. A. (2018). Automated analysis of language production in aphasia and right-hemisphere damage: Frequency and collocation strength. *Aphasiology, 32*(11). 1267–1283.